UNIVERSITY COLLEGE LONDON

Faculty of Mathematics and Physical Sciences

Department of Physics & Astronomy

# OPTICAL REMOTE SENSING OF MESOSCALE THERMOSPHERIC DYNAMICS ABOVE SVALBARD AND KIRUNA

Thesis submitted for the Degree of Doctor of Philosophy of the University of London

by

Amy Ronksley

Supervisors:

Dr Anasuya Aruliah

Prof Alan Aylward

Examiners:

Dr Gareth Chisham

Dr Nick Achilleos

February 25, 2016

*This thesis is dedicated to the household of Selwood Villa, Andrew, Patricia and Harold Ronksley, to whom I owe so much.*

I, Amy Marie Ronksley, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

# Abstract

Vertical winds are key in thermospheric dynamics and only until recently have the detectors been sensitive enough for them to be measured accurately. Two narrow field Fabry-Perot Interferometers (FPIs) are used as well as one state-of-the-art all-sky FPI, SCANning Doppler Imager (SCANDI), which is capable of simultaneous measurements across the sky at high spatial and temporal resolution. They measure the atomic oxygen 630nm emission line which peaks in brightness at 240km altitude in the upper thermosphere region. Emission intensities, line-of-sight wind speeds and neutral temperatures are obtained.

SCANDI's existing infrastructure has been developed based upon the requirement to upgrade the sky map to higher spatial resolution, for the onset of solar maximum. The calibration methods and data analysis are presented. The wind-fitting algoithm is shown for the new map trigonometry. This fitting is verified by producing climatological horizontal wind-fields in a dial plot format and cross-comparing with SuperDARN climatologies.

A statistical analysis of the vertical winds from 2002-2009 is presented leading to the possibility of 'black swan events' around midnight in the polar cap. These are events which are thought impossible but are, in reality, found to have a small finite chance of occurrence. An investigation into the mechanism of the generation of these events leads to the discovery of hydroxyl contamination in the Svalbard data set. A spectral simulation of the 630nm and the hydroxyl lines allows the determination of an emission intensity threshold of 40R (10R) below which the wind (temperature) values are significantly affected. The Svalbard data set is re-analysed excluding the contaminated data and a clean, more reasonable data set is presented with no black swan events.

A statistical study of the relationship between the vertical and horizontal components

of wind is presented showing the Burnside relationship is unsuitable for representing high-latitude winds. The CMAT2 atmosphere model data is used to assess which of Burnside et al's (1982) assumptions are violated.

The CUSPN campaign is presented showing the first results of the charged and neutral cusp region being simultaneously and independently measured using the EISCAT Svalbard Radar and the FPIs. Characteristic upwellings are observed concurrent with cusp precipitation and flux transfer events, which provides compelling evidence of high altitude ion-frictional heating.

# Acknowledgements

I would like to thank my supervisor Dr Anasuya Aruliah for providing invaluable guidance throughout my PhD, being a patient and understanding pastoral figure as well as my boss; without her help the materialisation of this thesis would not have been possible. Dr Ian McWhirter has been a huge help, frequently explaining the intricacies of the instrument mechanics and also offering up his own time in order to give valuable thesis advice. My second supervisor Prof. Alan Aylward at the Atmospheric Physics Laboratory has very generously given useful advice on my research even though he is retired. I would also like to thank Dr Iris Yiu, my PhD predecessor, for all her training and advice on the SCANDI data analysis after she finished. Thanks to Dr Timothy Spain who provided me with lots of supervision in my first two years.

I am grateful for the Natural Environment Research Council (NERC) for funding this PhD studentship. I am also grateful to the FPI/SCANDI designers and engineers, Dr Ian McWhirter and Dr Andy Charalambous, for creating the tools which provide my data. Thanks again to Dr Aruliah for the raw FPI data analysis which has provided me with a useable data-base. Thanks to Prof Dag Lorentzen (and his PhD student Xiangcai Chen), at the University Centre in Svalbard for supplying me with countless MSP data files from the Kjell Henriksen Observatory (KHO) on Svalbard, Norway. I am grateful to Dr Eoghan Griffin and Prof Mark Conde for the data-analysis routines inherited from them. Prof Mark Conde was also a great help with his insight into spectral contamination. His ex-PhD student, Dr Callum Anderson, very generously provided me with a great deal of his expertise on the SCANDI wind-fitting procedure (via email), even though he did not work in the field anymore. Thanks to Margit Dyrland for her help with the logistics of the CUSPN campaign as well as 'showing the APL a good time' in Svalbard. Thanks

to the University of Leicester (and David Barnes as the middle-man) for supplying me with a data-base of analysed SuperDARN data. Thanks to David Johnson for providing me with a set of CMAT2 data. EISCAT is an international

The entire KHO staff from UNIS, ESRANGE KEOPS facility and the Sodankylä Geophysical Observatory have been very generous with the logistics and maintenance of our instruments. Thanks to EISCAT, an international association supported by research organisations in China (CRIRP), Finland (SA), Japan (NIPR and STEL), Norway (NFR), Sweden (VR), and the United Kingdom (NERC), for providing invaluable supplementary data. Thanks to Prof Herb Carlson and Prof John Meriwether for their expertise on the CUSPN campaign. Thanks to the IMAGE magnetometer network collaboration between Finland, Germany, Norway, Poland, Russia and Sweden.

Many thanks also go to my family for their relentless support and generosity throughout my education; as well as Milo Coomber for his generosity and kindness throughout my PhD.

# Contents

This page was intentionally left blank

# List of Figures

This page was intentionally left blank

# Chapter 1

# Introduction

## 1.1 Sun-Earth Interaction

The solar terrestrial environment surrounding the Earth hosts a myriad of time-varying interacting processes involving energetic particles, waves, hot plasmas, magnetic and electric fields. This complex and dynamic system, particularly in the near-Earth region, is termed space weather. Space weather is conceptually similar to meteorological weather however it is less to do with the movement of bodies of water and more to do with plasma drifts and electromagnetic fields, as well as occurring hundreds of kilometres higher up in the atmosphere. The overlap between the two types of weather is due to their shared common drivers, the predominant one being the Sun and others including the Coriolis force and atmospheric gravity waves.

First and foremost, the Sun interacts with the Earth via solar radiation which takes approximately eight minutes to arrive. Then, there is the more direct interaction from the solar wind, the hot fast-moving plasma which radially emanates from the Sun. This physically connects to the Earth's magnetic field (geomagnetic field), and takes between 4-6 days to arrive depending on the speed of the solar wind.

The solar irradiation is responsible for the global-scale heating of the upper atmosphere driving the general circulation from the sunlit dayside region to the cooler night-side region (Roble, 1983). The radiation is also responsible for ionising the upper atmosphere creating the ionosphere which, when combined with the geomagnetic field, is associated

with the flow of currents and plasma drifts. This enables coupling between the neutral and charged atmosphere via collisional momentum and energy transfer between the ions and the neutrals.

The solar wind's interaction with Earth drives a magnetospheric plasma convection around the Earth and sets up the terrestrial current systems. If a particularly energetic solar event occurs, such as a coronal mass ejection (CME), the solar wind can disturb the geomagnetic field and initiate geomagnetic storms, i.e., substorms, which enhance the ionospheric currents and can result in beautiful intensified auroral displays at high latitudes.

### 1.1.1 The solar wind

The solar wind is a hot stream of fast-moving plasma (mostly protons and electrons) emitted radially outwards from the sun (Hundhausen, 1995). It is highly conductive and has a very low number density varying between $<1\text{cm}^{-3}$ and $20\text{cm}^{-3}$ at Earth's orbit (Rees, 1989). The typical speed of the wind is $400\text{kms}^{-1}$ (Gosling and Hansen, 1971) however can vary between $250\text{kms}^{-1}$ and $800\text{kms}^{-1}$. Occasionally a coronal mass ejection (CME) is released from the Sun whereby huge quantities of plasma ($\sim 10^{12}$kg) are explosively ejected from the active sun spots. This can increase the number density of the solar wind. For the most explosive events, plasma can be ejected at speeds of $>2000\text{kms}^{-1}$ (Gopalswamy et al, 2004).

The interplanetary magnetic field (IMF) is the field that the solar wind carries with it; the IMF is 'frozen in' to the solar wind. The frozen in field is a useful simplification which aids the understanding of the gross solar wind-magnetosphere interaction. It is an idealised condition where the plasma is assumed infinitely conductive and collisionless which means the particles cannot cross over the field lines and so the plasma can be visualised as 'carrying' a magnetic field along with it. If the plasma had a finite conductivity the field lines would diffuse through the plasma. Technically, this is the reality for the solar wind as the diffusion time-scale ($\tau_d$) is given by (Priest, 1995):

$$\tau_d \approx \mu_0 \sigma L^2 \tag{1.1}$$

where $\mu_0$ is the permeability of free space ($4\pi\times\ 10^{-7}\text{WbA}^{-1}\text{m}^{-1}$), $\sigma$ is the conductivity ($\text{Sm}^{-1}$) and L is the characteristic scale length (m). However as the diffusion time-scale

is proportionate to the conductivity and the solar wind conductivity is so large, the time-scale would be very long relative to the bulk flow. Therefore, the frozen-in assumption is a reasonable one, both for the solar wind and the magnetospheric plasma.

The magnitude of the IMF at the orbit of the Earth is of the order $10nT$. One of the co-ordinate systems used to describe the IMF is the GSM (Geocentric Solar Magnetospheric). It is a three-dimensional system with respect to the Earth - Bx (positive towards the Sun along the Earth-Sun direction), By (dawn to dusk) and Bz (north to south along Earth's magnetic dipole). The By and Bz components influence the strength of the interaction between the solar wind and the geomagnetic field.

### 1.1.2 The magnetosphere

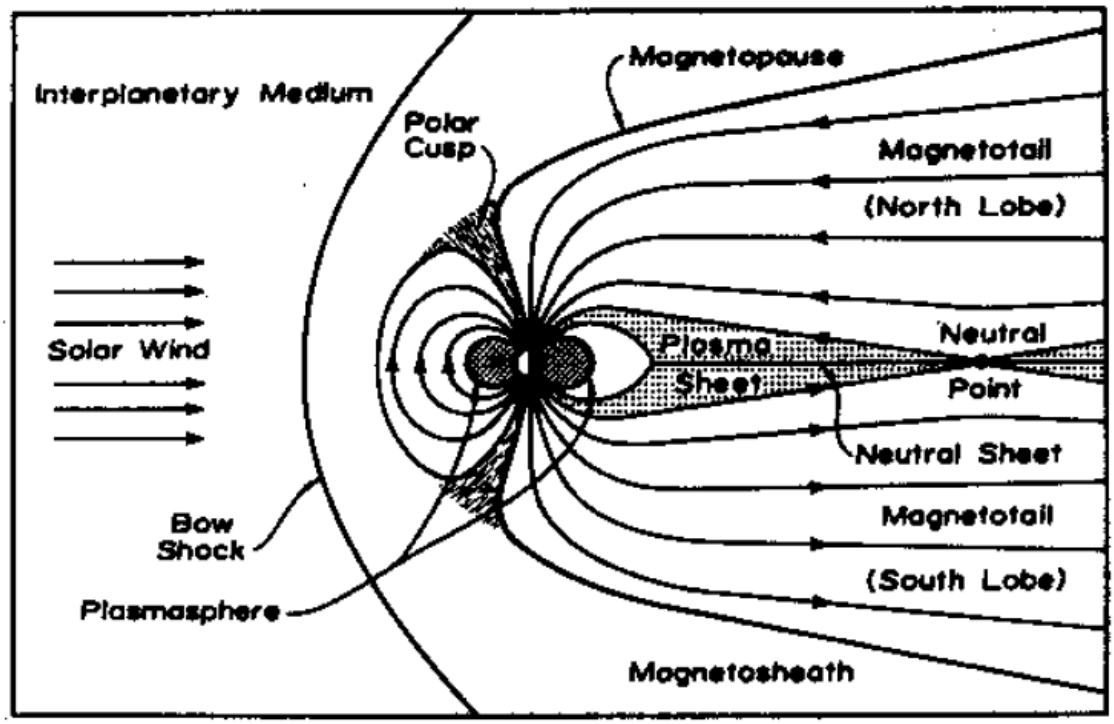


Figure 1.1: Schematic noon-midnight meridian cross-section of the terrestrial magnetosphere (Russell, 1972).

The magnetosphere is a protective bubble surrounding the Earth formed by the interaction between the geomagnetic field and the IMF. It is key to life on Earth by firstly acting as a shield to prevent the atmosphere from being stripped away by the pressure of the solar wind. Secondly, it acts to guide harmful charged particles around the Earth so they generally do not reach the surface and endanger the biosphere. Without the magne-

tosphere, the human race would be subject to the same radiation threats as interplantary astraunauts. The most serious threat is thought to come from, the sporadic, solar proton events (SPEs). If present in space at the time of an SPE, an astronaut would be bombarded with a large dose of harmful proton radiation, which is likely to cause acute radiation sickness and can be rapidly fatal (Seed, 2011). There are further health concerns from the bombardment of galactic cosmic rays which cause carcinogenesis of cells (Cucinotta and Durante, 2006) and an acceleration of neurological degeneration (Vazquez, 1998). For astonauts on month long missions (a typical mission length) experiencing an SPE is unlikely and exposure to GCRs is relatively short, however over a lifetime on Earth, without the protection of a magnetosphere, the likelihood of experiencing an SPE is significantly increased and the exposure to GCRs will be persistent. Global life expectancy would no doubt decrease significantly.

A schematic drawing of the magnetosphere is shown in figure 1.1 displaying a cross-section of the terrestrial magnetosphere along the noon-midnight meridian. The magnetospheric magnetic field has large departures from the typical dipole shape magnetic field that is expected if the Earth was completely isolated. The dynamic pressure of the solar wind compresses the geomagnetic field on the dayside and on the nightside the field lines are stretched many Earth radii downstream forming the magnetotail. On the dayside a 'bow shock' is formed where the solar wind initially interacts with the geomagnetic field and is abruptly diverted around the magnetospheric cavity. Slightly Earthward of this feature is the magnetopause which marks the official boundary between the plasma of the solar wind and the magnetosphere. This typically sits roughly 10 Earth radii ($R_e$) away from the earth, at its subsolar location, although it can vary by 1-2 $R_e$ depending on the rate of dayside magnetic reconnection and the solar wind dynamic pressure.

There are two types of magnetospheric configurations: closed and open. These are displayed in figure 1.2 a) and b) as diagrams of noon-midnight meridian cross-sections of the magnetosphere within the solar wind field which flows from left to right. The set-up of the configuration is controlled by whether the IMF is orientated northward ($B_z > 0$) or southward ($B_z < 0$). The solar wind's interaction with the geomagnetic field is strongest when there is a southward orientation (see subplot b)). This permits an 'open' magnetosphere where the geomagnetic field directly merges with the IMF (field lines labelled 'C') allowing direct access for solar wind particles into the magnetosphere, thus the term 'open' field lines. A northward orientation of the IMF ($B_z > 0$), however,

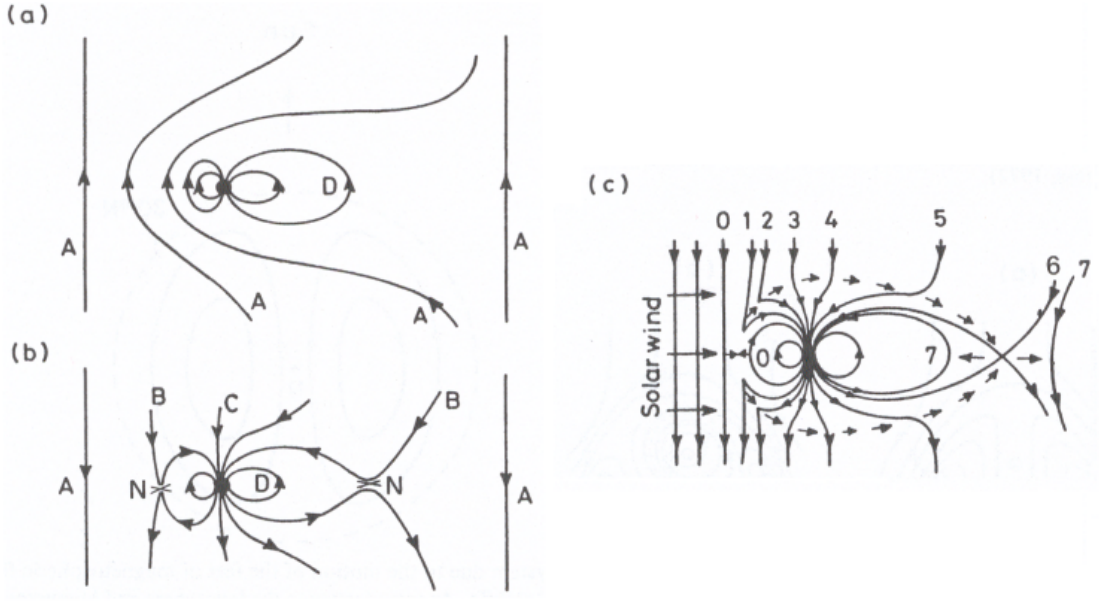


Figure 1.2: Terrestrial and solar wind magnetic field configuration for, a) northward IMF and, b) a southward IMF and, c) the convection pattern over the polar cap (Dungey cycle) for a southward IMF. A: interplanetary magnetic field lines, B: IMF lines reconnecting to geomagnetic field line, C: open, D: closed and N: neutral points or reconnection sites, i.e., dayside reconnection point and near-Earth neutral line (Hargreaves, 1979).

creates a 'closed' magnetosphere (see subplot a)) where there are no shared field lines between the solar wind and geomagnetic field, thus all of the Earth's field lines are closed (labelled as 'D').

In the special case of the open magnetosphere a magnetic flux tube convection about the Earth is established termed the Dungey cycle after its original discoverer. This is shown in figure 1.2 c) where the different stages of the cycle are labelled from 0 to 7. The Dungey cycle is due to antiparallel orientations (or at least orientations with relatively high magnetic field shear) of the southward IMF and the northward geomagnetic field (see stage 0) creating conditions which promote a process called magnetic reconnection. This process involves anti-parallel field lines breaking at a location where the field changes abruptly across relatively short spatial scales and rejoining to each other creating two new field lines. These can quickly convect away from the site of reconnection by a sudden release of energy stored in the tension of the magnetic field lines. This process is shown in figure 1.3.

There are two areas in the open magnetosphere where the conditions are right for reconnection. These are both labelled as 'N' for 'neutal point' (where locally B=0) on

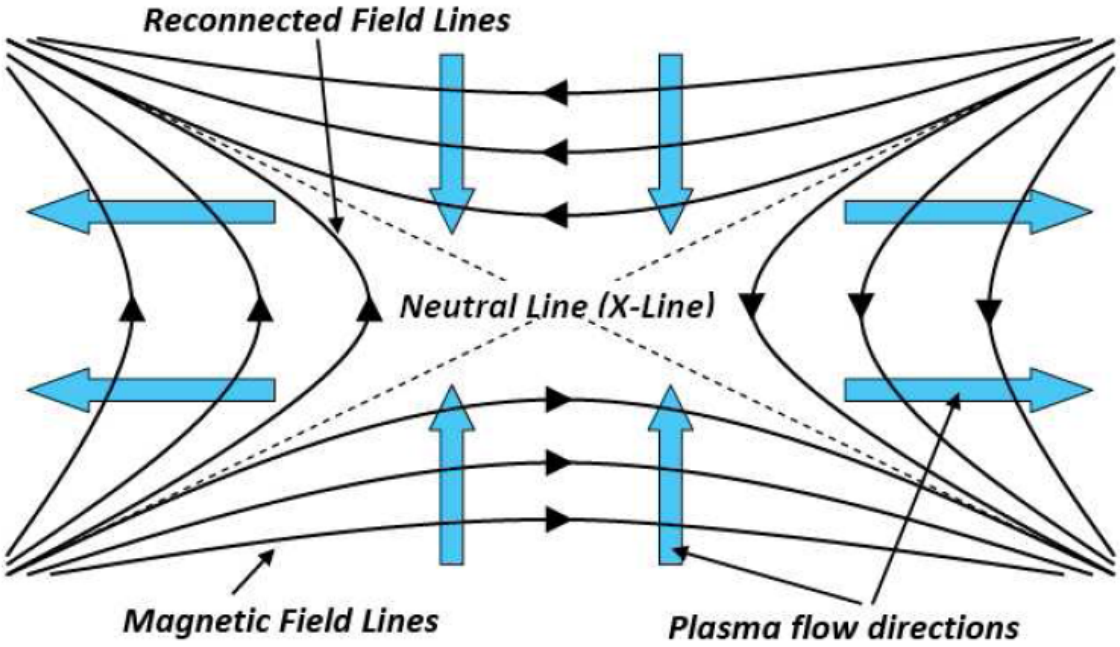


Figure 1.3: A magnetic reconnection diagram showing antiparallel magnetic field lines from two different locations (lines at the top and bottom of figure) reconnecting to one another (at the 'Neutral Line' point in the centre) in order to form new field lines (at left and right of figure) which convect away from the reconnection site (Anderson, 2011).

subplot b). The first one is at the front of the magnetopause, i.e., the dayside reconnection point, and the second one is at the tail end of the magnetosphere, i.e., the near Earth neutral line. The combination of these events drive a convection of magnetised plasma around the Earth (the Dungey cycle). The dayside reconnection event (stages 0-1) creates two 'open' field lines with their footprints originating in the Earth's core and the other ends joined to the solar wind (this is part of the polar cusp region in figure 1.1). As the solar wind continues downstream it effectively drags the open field lines anti-sunward over the poles (stages 2-4) back to the tail of the magnetosphere (stage 5) where they meet again and, in a simple steady-state picture, the second reconnection event occurs (stages 6-7). The magnetic field lines are once again closed and convect back around opposite equatorward sides of the polar cap, i.e., the dawn and dusk flanks of the magnetosphere, until they meet once again at the dayside reconnection point and the convection cycle repeats itself.

In this open magnetosphere configuration, there is a careful balance between the two reconnection events. The dayside reconnection feeds open flux into the magnetosphere and the nightside process removes it, converting it back to closed flux. If the dayside

reconnection rate exceeds nightside, then open flux loading occurs which may eventually be dissipated explosively via a dramatic episode of nightside reconnection, i.e., geomagnetic substorms. When geomagnetic substorms, of typical duration of a few hours, occur persistently for a few days the disturbance can be classified as a full geomagnetic storm.

In the other magnetospheric configuration (shown in figure 1.2 a)), i.e., in the case of a closed magnetosphere ($B_z > 0$), the Dungey cycle is not established. This is due to there being no dayside reconnection. At the dayside subsolar location, the northward IMF field lines (labelled 'A') run parallel in the same direction as the closed geomagnetic field lines, hence, no reconnection occurs here. However, there is a small amount of reconnection promoted on the nightside of the magnetorsphere. As the solar wind flows downstream (beyond the dayside subsolar point), the field lines (A) drape around the top and bottom of the magnetospheric cavity and across its entire length (tailward). The draped field lines become antiparallel with the tailside closed geomagnetic field lines (D). Here, reconnection occurs on both the top and bottom, creating a new closed field line (which now loops round the dayside) and, also creating a new IMF field line (A) on the night-side (similar to the original IMF field lines (A) on the dayside). The rate of reconnection in this case, is far smaller than in the case of an open magnetosphere.

On time scales I am concerned with (>1sec), the convecting magnetic field lines are considered electrostatic equipotentials. This is due to the fact that the conductivities along them are so high that any slight charge imbalance would be immediately neutralised by a field-aligned current, hence maintaining an apparent zero electric potential gradient This condition may be violated in the so-called auroral acceleration region. Within this region, the Cluster spacecraft collected data from 4000 and 7000km altitude along the same magnetic field line to show a significant potential drop of 500V (Forsythe et al, 2012).

An implication of this phenomenon is that electrostatic equipotential surfaces are traced out by these convecting field lines which map down in altitude from the magnetosphere to the upper ionosphere. Due to the near-collisionless plasma in the F region, the frozen-in assumption is appropriate for describing the plasma flow. This plasma motion is termed $\vec{E} \times \vec{B}$ drift which defines the direction of horizontal flow in the upper thermosphere, i.e., perpendicular to both the electric and magnetic field directions, hence along the equipotential surfaces. For IMF $B_z < 0$, the motion is antisunward over the poles creating what is named 'the cross-polar cap jet' and the flow returns around the sides of

the polar cap creating two convection cells. For IMF $B_z$ >0, the flow is less predictable and there can be multiple smaller convection cells set-up. Figure 1.4 shows an average of SuperDARN radar (see section 1.7.3) data (Chisham et al, 2007) from 2002 (plot produced by myself) on a 200x200 magnetic latitude - magnetic local time grid (MLAT-MLT), i.e., a bird's-eye view of the magnetic pole with the sun situated at infinity off the top of the diagram. The black vectors represent the plasma drift, the blue lines show the electrostatic equipotential lines and the red circle shows Svalbard's (one of UCL's observing sites) path around the magnetic pole at ∼75° latitude.

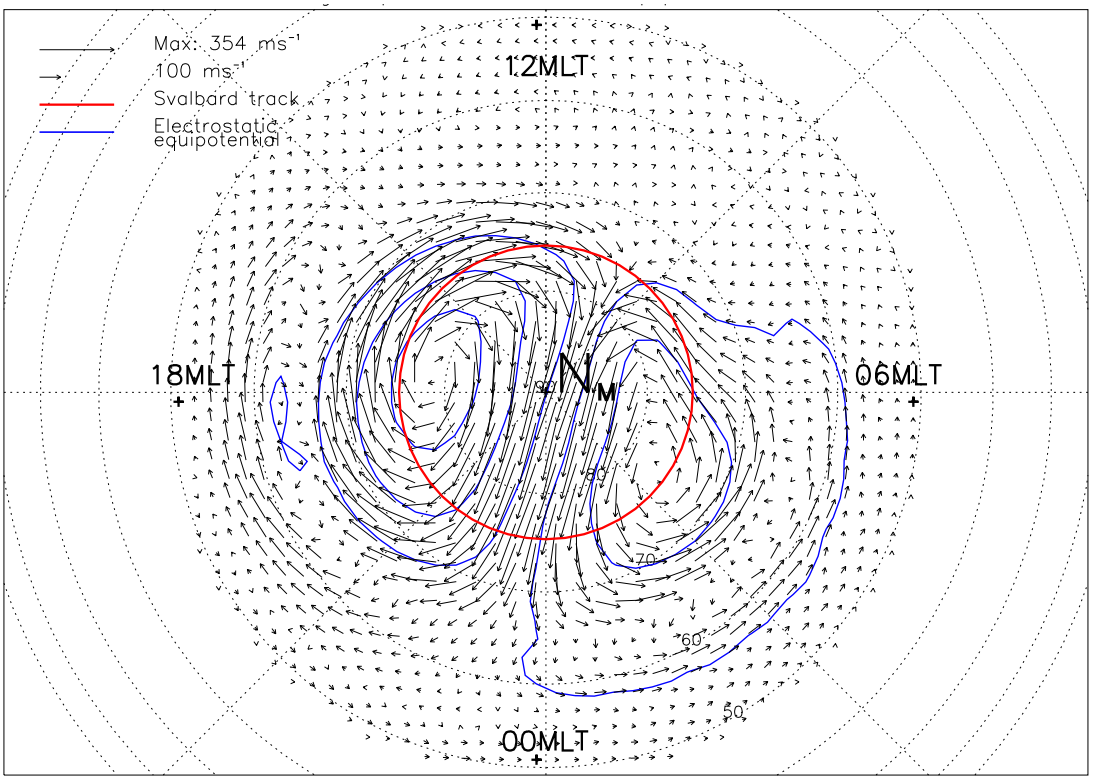


Figure 1.4: SuperDARN plasma drift vectors as black arrows (calculated from electrostatic potential data by myself with Ruohoniemi and Baker (2007) method and plotted by myself) averaged over entire year of 2002 plotted on an MLAT-MLT grid. The electrostatic equipotentials are in blue and Svalbard's track around the magnetic pole is in red. Data courtesy of University of Leicester (Chisham et al, 2007) and the SuperDARN collaboration (see section 1.7.3).

The cross-polar cap jet is on open field lines, following dayside reconnection, whereas the sunward return flow is carried on the closed field lines after nightside reconnection. The boundary between open and closed magnetic field lines separates these antisunward and sunward flows and manifests itself in the atmosphere as the visible light shows of the

auroral oval.

As the two magnetospheric reconnection processes are also mechanisms which accelerate charged particles into the atmosphere, these create aurora. The dayside reconnection is associated with flux transfer events (FTEs) and is responsible for the cusp aurora (see 'Polar Cusp' region in figure 1.1), whereas the tailside reconnection creates part of the nightside aurora. The cusp aurora occurs at a higher altitude than the night-side aurora due to the day-side magnetic field lines' smaller magnetic tension force relative to the nightside's. Therefore, the dayside reconnection supplies less kinetic energy to accelerate the particles (mostly electrons and protons) causing them to penetrate not so deeply into the atmosphere (∼240km altitude). These particles are called 'soft' precipitation with typical energies of ∼hundreds eV. However, the tailside magnetic tension is enormous, so upon a reconnection event, the field lines act as a 'great elastic band' catapulting the plasma Earthward with vast amounts of energy so that the charged particles can penetrate the atmosphere further down to lower altitudes. This is called 'hard precipitation' with typical energies >keV (typically at ∼110km altitude). Due to the chemistry prevailing at this altitude, this hard precipitation creates a green airglow which is the most common observational signature of aurora, whereas the soft precipitation creates a red glow which is fainter in the sky and so harder to observe with the naked eye.

### 1.1.3 Ionosphere

The ionosphere is the charged constituent of the atmosphere distributed from 60-1000km altitude. It accounts for <1% of the entire mass of the upper atmosphere, however, it is still paramount in the dynamics of this region due to its sensitivity to changes in geomagnetic forcings which manifest in the ionosphere as energetic currents and plasma flows. The ionosphere's electrical properties allow the reflection of upward-propagating radio waves which allows radio communication across the globe and also radar measurements of atmospheric parameters. The ionosphere is formed predominantly by ionisation of the major neutral species, $O$, $N_2$ and $O_2$, by solar EUV and X-ray radiation. However, at high latitudes, there is additional ionisation provided by magnetospheric energetic particles precipitating into the atmosphere and being accelerated by electric fields (parallel to the magnetic field). The basic solar photoionisation reactions are:

$$N_2 + h\nu \rightarrow N_2^+ + e^- \tag{1.2}$$

$$O_2 + h\nu \rightarrow O_2^+ + e^- \tag{1.3}$$

$$O + h\nu \rightarrow O^+ + e^- \tag{1.4}$$

Photons with sufficient energy may act to dissociate as well as ionise in these reactions:

$$N_2 + h\nu \rightarrow N^+ + N + e^- \tag{1.5}$$

$$O_2 + h\nu \rightarrow O^+ + O + e^- \tag{1.6}$$

Precipitating electron impact may also act to ionise, dissociate and excite molecular species as follows:

$$N_2 + e_p^- \rightarrow N_2^+ + e_p^- + e_s^- \tag{1.7}$$

$$O_2 + e_p^- \rightarrow O_2^+ + e_p^- + e_s^- \tag{1.8}$$

$$N_2 + e_p^- \rightarrow N^+ + N + e_p^- + e_s^- \tag{1.9}$$

$$O_2 + e_p^- \rightarrow O^+ + O + e_p^- + e_s^- \tag{1.10}$$

$$N_2 + e_p^- \rightarrow N^* + N + e_p^- \tag{1.11}$$

$$O_2 + e_p^- \rightarrow O^* + O + e_p^- \tag{1.12}$$

where an asterisk denotes an excited species, $e_p^-$ represents a primary incoming electron and $e_s^-$ a secondary electron expelled from the molecule.

There are three distinct regions of ionisation in the atmosphere called the D, E and F (F1 and F2) layers in order of altitude from low to high. The heights and intensity of these regions depend on the rate of recombination and ionisation, i.e., ionospheric loss and production. At night-time only the F layer remains ionised while the E and D regions almost disappear completely. This is attributed to the lower layers, firstly, being strongly dependent on solar radiation and, secondly, their rate of recombination is rapid due to the increased neutral and ion density causing an increased collision frequency. However in the day-time, all layers are significantly enhanced by photo-ionisation; the F layer even gains an additional layer at a slightly lower altitude named the F1 layer, with the original F layer becoming the F2 layer.

The D layer lies between 60-90km altitude and is formed by ionisation of nitric ox-

ide ($NO$) via Lyman-alpha radiation from hydrogen (121.5nm), as well as ionisation of molecular nitrogen and oxygen by hard X-rays(1-10Å). At night-time residual amounts of ionisation are formed from cosmic rays. The E layer peaks at 105-110km altitude and is formed by ionisation of molecular oxygen by soft X-ray (1-10nm) or far UV predominantly creating $NO^+$ and $O_2^+$ and also plenty of $O^+$. These ion reaction products can quickly recombine with free electrons at these altitudes.

The F layer is the most densely ionised region of the ionosphere. The abundance of atomic oxygen ions exceeds that of $NO^+$ and $O_2^+$. F1 peaks at 160-180km altitude. It is formed from chain reactions triggered by the absorption of solar EUV, creating ions of $NO^+$ and $O_2^+$ with intermediate products of $N_2^+$, $O^+$, $He^+$ and $N^+$. The F layer is also subject to charge exchange processes between molecular oxygen, atomic oxygen and ionic oxygen, supplying surplus atomic oxygen. F2 peaks at 200-400km altitude and is formed predominantly by dissociative recombination reactions (see section 1.2.2). Solar photoionisation is not an important process here relatively speaking, since the solar radiation deposits more energy for ionisation at lower altitude regions. More information on the chemical processes and composition is given in section 1.2.2.

In the near-collisionless F region of the ionosphere plasma drift is largely controlled by the $\vec{E} \times \vec{B}$ drift. At lower altitudes in the atmosphere the neutral density increases so the number of ion neutral collisions per unit time increases, providing a resistive friction on the ions known as ion drag. This rotationally accelerates the ions, however, it has a negligible effect on the electrons which continue along their paths relatively unimpeded. The separation in the flow of the ions and electrons means they can carry a current. This is the reason for the peak of the horizontal current density lying in the E region and hence the reason for the peak of Ohmic Joule heating (heating mechanism associated with currents flowing in a medium of finite conductivity, explained more in 1.3.2) also lying here. The current density ($\vec{j}$) at a given height in the ionosphere is given by Brekke (1997):

$$\vec{j} = n \cdot e \cdot (\vec{v_i} - \vec{v_e}) \tag{1.13}$$

where $\vec{v_i}$ and $\vec{v_e}$ are the ion and electron drifts ($\mathrm{ms}^{-1}$), $n$ is electron density ($\mathrm{m}^{-3}$) and $e$ is the electron charge ($1.6{\times}10^{-19}$C).

### 1.1.4 Thermosphere

The thermosphere is the neutral constituent of the upper atmosphere, spanning from ∼80-600km altitude. It is generally considered an inactive medium compared to its charged counterpart (the ionosphere) due to its highly viscous nature and, above 250km altitude, its relatively small altitudinal velocity variation. The neutrals here make up more than 99% of the mass of the upper atmosphere so its impact cannot be overlooked. The thermosphere's viscosity increases with altitude due to the increasingly large mean free path and kinematic temperature (until the isotherm is reached at ∼300km altitude). The temperature profile is a result of the balance between the absorption of solar radiation and the downward molecular conduction of heat. Even though Joule heating peaks at ∼150km altitude, the temperature maximum is not reached until ∼300km altitude due to the lower neutral density at higher altitudes meaning that there is a larger Joule heating per unit mass (Deng et al, 2011). Regardless of the positive temperature gradient and previous misconceptions, work has revealed that the thermosphere is far from being a passive medium and is, in fact, an important sink of magnetospheric energy and source of ion-frictional heating (Oyama et al, 2004).

The thermosphere has relatively large scales of motion, for example, at 300km altitude the mean free path is about 1km, the mean time interval between collisions is about 1 second (Aruliah, 1991) and the shortest time scale (in wind and temperature) is assumed to be ∼1hr. However, these results have been challenged in recent years with the development of higher resolution measurements by Fabry Perot Interferometers (FPIs) and much mesoscale variability has been revealed (Aruliah et al, 2004; Kosch et al, 2011; Anderson et al, 2009, 2012). The time scales have also been investigated using FPIs, e.g., Ford et al (2007) used 15-second resolution FPI wind and temperature measurements to determine a minimum time scale of thermospheric variability of a few minutes. This variability is paramount in deriving accurate estimates of the Joule heating rates when including the neutral wind dynamo (Aruliah et al, 2004; Thayer et al, 1998). Current atmospheric models cannot simulate this small scale variability due to their coarse grid sizes of a few degrees latitude and longitude, which approximates to several hundred kilometres.

The thermospheric circulation is a result of a combination of different forcings. The diurnal solar heating creates a day-to-night pressure gradient which drives the global circulation. This is complicated by the Coriolis force and, at high latitudes, coupling with

the ionosphere via ion drag, creating a strong cross-polar-cap, day-to-night jet. This is explained in more detail in the section 1.2.3.

Historically, the thermosphere has not been as popular to study as the ionosphere, due to the difficulties in observing this region and its apparent passivity. Apart from rockets' and satellites' in situ measurements, all other observations are via remote optical ground-based instruments which rely on both clear skies and darkness which, for the polar regions, leaves only winter time measurements. On the other hand, the ionosphere is typically observed via radar and can be observed 24 hours a day as there is always some ionisation present. Furthermore, the thermosphere is subject to a 'geomagnetic history' whereby the geomagnetic conditions of the previous 3-6 hours affects the neutral wind as well as the current conditions, so should be accounted for when calculating Joule heating and momentum transfer (Aruliah and Müller-Wodarg, 1999).

### 1.1.5 MIT system

The magnetosphere-ionosphere-thermosphere (MIT) system describes the tight coupling between the three constituents of the terrestrial environment. This is largely achieved via atmospheric dynamos, ion drag and field aligned currents supported by particles precipitating into the ionosphere from the magnetosphere, feeding the auroral electrojet. Enhanced ionospheric density can affect the neutral composition via chemical reactions (see section 1.7) and can also effect the heating of the thermosphere, thus, moving neutral species to new altitudes. The thermosphere can also be a great heat sink for the ionosphere and magnetosphere having been reported to receive ∼60% of the magnetosphere's energy input (Killeen et al, 1985). The thermosphere is also known to feed energy back into the ionosphere via the neutral wind dynamo and ion-drag.

The magnetospheric dynamo, -($\vec{V} \times \vec{B}$), is the convective electric field of a collisionless plasma, from the plasma drift ($\vec{V}$) across geomagnetic field lines ($\vec{B}$). The neutral wind dynamo, ($\vec{U} \times \vec{B}$), arises when the neutral wind ($\vec{U}$) forces ions across field lines hence, in effect, generating an electric field. Both dynamos produce currents and feed energy back into the ionosphere however from two opposite sides of the MIT system: the magnetosphere and the thermosphere.

The upper ionospheric plasma convection is largely influenced by the clock angle of the IMF. This is the ratio between By and Bz components (GSM coordinate system) which affects the degree of magnetosphere-ionosphere coupling.

The magnetosphere is coupled to the ionosphere via the Birkeland current system, a system of field-aligned currents (FACs), which flow into and out of the high latitude ionosphere along the magnetic field lines and close in the tail current. The magnitude of the coupling is further controlled by the amplitude of the IMF Bz negative component and the IMF By component. Some FACs enter the ionosphere along the auroral oval and close via the auroral electrojet currents which flow perpendicular to the magnetic field lines. These auroral FACs allow the transport of highly energetic magnetotail electrons into the ionosphere after nightside reconnection, creating aurora.

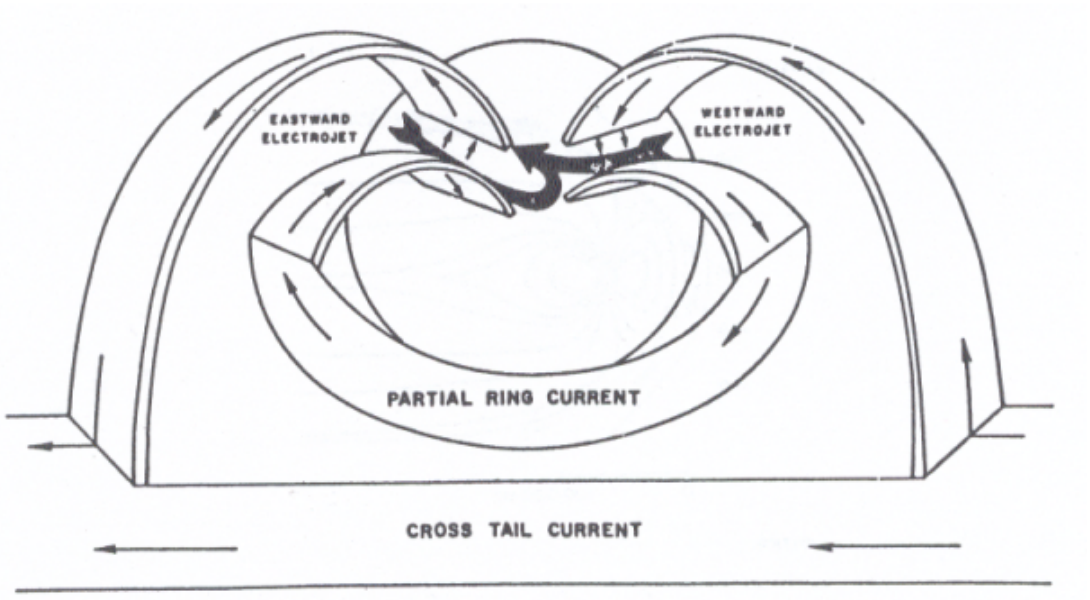


Figure 1.5: Birkeland current system showing field-aligned currents being closed in the ionosphere via the auroral electrojet (Hargreaves, 1992).

The FAC system consists of a current flowing in from the cross tail current to the post-midnight region of the lower ionosphere at the top of the auroral oval. It flows equatorward horizontally and then at this lower latitude flows upwards out of the ionosphere towards the partial ring current, closer to Earth than the cross tail current is. This ring current flows in the westward direction crossing magnetic midnight before flowing back down the field lines reaching the lower-latitude ionosphere this time on the pre-midnight side. The current then flows horizontally poleward to a higher latitude at the top of the auroral oval, where it exits the ionosphere once again flowing up the field line out to join the dusk side of the tail current. In the ionosphere at the top of the auroral oval the currents coming in at post-midnight and exiting at pre-midnight are connected via the westward electrojet.

This situation is reversed at the equatorward side of the auroral oval where the current enters the ionosphere on the pre-midnight side and exits on the postmidnight side, these systems being connected via the eastward electrojet. The electrojets are part of a closed magnetospheric current system.

Another complementary and more mechanical explanation of the existence of the auroral electrojets is that they are a result of the increased neutral density in the lower atmosphere. In the upper ionosphere the flow is controlled by the $\vec{E} \times \vec{B}$ drift and follows the convection pattern, diverging sunward around the flanks of the polar cap after tail-side reconnection. The increased neutral density in the lower thermosphere rotationally accelerates the ions in the opposite direction from the general drift causing the ions to separate from the electrons, hence, creating currents. These are directed back antisunwards around the flanks of the auroral oval converging around magnetic midnight. These are the eastward electrojet in the pre-midnight sector and the westward electrojet in the post-midnight sector.

Ion drag is the term given to this collisional resistive friction that the neutrals provide to the ions and results in an increase in the ion temperature, i.e., ion frictional heating. This is a direct coupling of the ionosphere and thermosphere. The magnetosphere's role in this mechanism also cannot be left out of this picture as it provides the electric field (mapped down from the magnetosphere) which acts on the charged particles and partly drives their motion, i.e., plasma drift. Without this plasma drift, there would be no collisions between the ions and neutrals which, as has been discussed, is responsible for the ion-frictional heating in the F region and the currents in the E region (which result in Ohmic Joule heating (see section 1.3.2)). The complete picture is one of a magnetosphere-ionosphere-thermosphere (MIT) system.

## 1.2 Introduction to Earth's Upper Atmosphere

Earth's atmosphere can be defined in several ways, including: temperature, neutral density and ion density with altitude. The height profiles of these different parameters are shown in figure 1.6, where you can see the different regions of the atmosphere, as well as the magnetosphere, labelled for the temperature and ion density. The neutral density is shown to have an exponential decrease with altitude however the particular composition varies and is discussed in section 1.2.2. The two peaks in the ion density profile correspond to

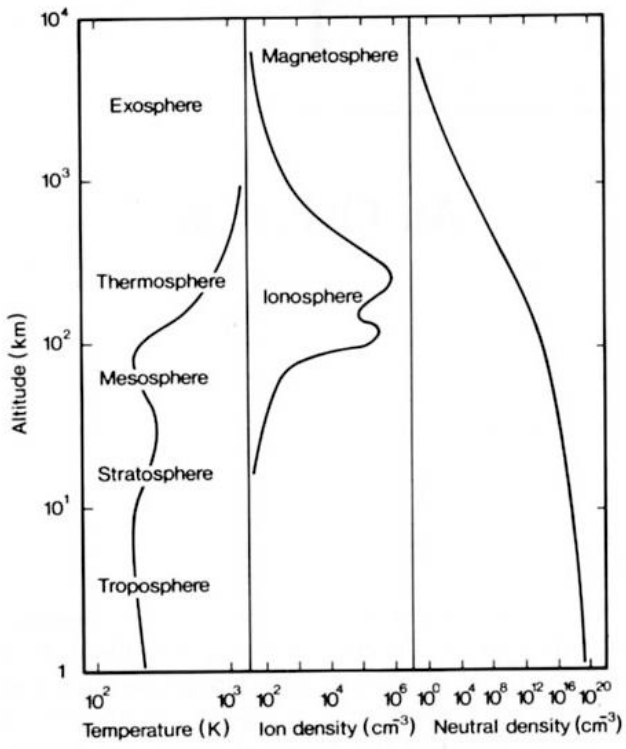


Figure 1.6: The temperature, ion density and neutral density altitude profiles of the Earth's atmosphere (Rees, 1989).

the E and F ionospheric layers discussed previously in 1.1.3. The most commonly used atmospheric region terminology describes the temperature gradients and these regions are labelled in the temperature profile diagram. Homogeneity is an additional useful means of characterisation of the atmosphere that has not been labelled on the diagram but is discussed further on.

Temperature gradients are the most common way to define the different layers of the atmosphere. The troposphere is the lowest level, extending from the Earth's surface to 10-12km. It contains three quarters of the mass of the whole atmosphere including almost all water vapour creating an environment for life, weather and aviation travel. It is heated from re-radiation of solar energy from the surface of the Earth, thus having a negative temperature gradient with altitude. This promotes convection and vertical mixing encouraging weather system formation. It is bounded by the tropopause which varies in height with season and latitude, from 4km at winter polar regions to 18km in summer equatorial regions. The next layer of the atmosphere is the more stable stratosphere where the temperature slowly rises from 210K up to 270K at the stratopause at an altitude of

∼50km. The stratospheric ozone absorbs solar UV radiation, heating the atmosphere and creating a positive temperature gradient, i.e., stable atmospheric conditions. This restricts weather production and cloud formation except for the occasional nacreous cloud (i.e., polar stratospheric clouds) observed at the lowest, coolest region. The next layer is the mesosphere which is the region where meteors burn up and noctilucent clouds exist. It is marked by another negative temperature gradient reaching a minimum of ∼180K at ∼80-90km altitude, i.e., the coldest part of the planet. This is due to molecular radiative cooling by $CO_2$ and $NO$ emissions.

The penultimate atmospheric layer is the thermosphere, the dynamics of which are the central study in this thesis. It is characterised again by a positive temperature gradient from heating mainly by solar EUV and X-ray absorption. It is the region which hosts the ionospheric layers and currents, as well as spanning the altitude range in which satellites orbit the Earth. Due to the lower air density and the low rate of re-radiation it reaches an unprecedentedly high temperature between ∼600-1600K (depending on solar cycle and seasonal conditions) at approximately ∼300km altitude. The final layer of the atmosphere lies above ∼600km altitude, and is known as the exosphere, where the particles move in ballistic trajectories with relatively small collision rates.

The homogeneity of the atmosphere is used to define two regions: the heterosphere and the homosphere, separated by the turbopause which lies at roughly 100km altitude. The homosphere is the region where all chemical constituents are mixed uniformly due to turbulence and the heterosphere is the region where the different chemical species are separated by molecular diffusion. This is due to the ratio of the mean free path to spatial mixing scale. This is large in the upper atmosphere, allowing diffusive separation to create a mean molecular weight gradient with the heavier thermospheric species such as oxygen and nitrogen more abundant at lower altitudes. However, the ratio is relatively small in the lower atmosphere, creating a well mixed, homogeneous gas.

### 1.2.1 General properties

To describe the structure of the neutral atmosphere, four parameters are required: the pressure, temperature, density and the composition. These are all interlinked within the ideal gas law:

$$P = k\frac{\rho}{M}T \tag{1.14}$$

where P is the pressure of the gas in Pascals (Pa), M is the mean molecular mass (kg), $\rho$ is the density (kg $m^{-3}$), k is the Boltzmann constant ($1.38\times10^{-23}m^2kg\ s^{-2}K^{-1}$) and T is the temperature of the gas in Kelvin (K). Another reasonable approximation for the atmosphere, at least on a global scale, is hydrostatic equilibrium:

$$\frac{dP}{dz} = -\rho g \tag{1.15}$$

This describes how the downwards force of gravity balances the upwards pressure gradient force, which mostly prevents the atmosphere from diffusing upwards and escaping the Earth. There will, however, be a small loss in atmospheric mass via diffusion, as some fraction of the exospheric particles' maxwellian distributions will have energies corresponding to particle velocities higher than the escape velocity for the Earth. Thus, these particles (a negligible fraction of the entire atmosphere) will escape the Earth's gravitational field.

The scale height is another useful quantity to describe the atmosphere as it is the vertical range over which the atmosphere will have similar characteristics and behaviour, therefore the physical parameters and processes can be assumed constant. More explicitly, the scale height is the vertical distance in which the pressure changes by a factor of $e$ (the base of natural logarithms). The scale height equation can be defined first by rearranging equations 1.14 and 1.15 with P as the subject and equating the two gives:

$$\frac{dP}{P} = -\frac{Mg}{kT}dz \tag{1.16}$$

and integrating with respect to z gives the equation describing the exponential decrease in pressure with altitude, the law of atmospheres:

$$P(z) = P_0 e^{-\frac{Mg}{kT}(z-z_0)} = P_0 e^{-\frac{(z-z_0)}{H}} \tag{1.17}$$

where the scale height (H) is assumed constant and is defined by:

$$H = \frac{kT}{Mg} \tag{1.18}$$

In the lower atmosphere the scale height is roughly constant ($\sim$7km) due to the turbulent mixing and the relatively uniform temperature with altitude. However in the upper atmosphere, specifically above the turbopause ($\sim$100km altitude) there is a substantial

variation in the scale height due to a dramatic increase in temperature and due to the diffusive separation of the chemical constituents causing the mean molecular mass to decrease with altitude. Referring to the previous equation, in the thermosphere the scale height must therefore increase with height, known to reach ∼50km in the F2 region.

The excited atomic oxygen ($O(^1D)$) subpopulation is in local thermodynamic equilibrium with the surrounding neutral gas at this altitude. This is because, the radiative lifetime of atomic oxygen, $O(^1D)$, is ∼110 seconds, whereas, the mean collision frequency at 250km altitude is ∼0.3 seconds. Therefore, the emitting oxygen atoms have many collisions and become thermalised before emission. Thus, the radiation is representative of the entire population of the neutral gas from the altitude of the emission.

Adiabatic processes are crucial in understanding the dynamics of the upper atmosphere. They are reversible processes which do not transfer of heat from system to surroundings; their energy transfer is achieved through work. Adiabatic heating is caused by an increase in the pressure of the air parcel by work being applied on it by its surroundings. Conversely, adiabatic cooling is where a parcel applies work on its surroundings due to a decrease in the surrounding pressure. When a parcel of air is forced up to a higher altitude, there will be decreased pressure on it so that the parcel expands in volume and so its temperature decreases, thus decreasing the internal energy. The rate at which the parcel heats or cools is used to derive the adiabatic lapse rate of the atmosphere:

$$\Gamma_{adiabatic} = -\frac{dT_{parcel}}{dz} \tag{1.19}$$

where $dz$ is the distance moved vertically and $dT_{parcel}$ is the temperature of the air parcel. This means a parcel after moving from altitude $z_0$ to $z_1$ ($=z_0+dz$) will have the temperature given by:

$$T_{z_1} = T_{z_0} + \Gamma_{adiabatic} dz \tag{1.20}$$

The actual atmospheric vertical temperature gradient, or the ambient lapse rate, has a large influence on the stability of the atmosphere, i.e., the resistance to vertical motions. Let us define the ambient lapse rate for an atmospheric temperature change $dT_{atm}$ as:

$$\Gamma_{ambient} = -\frac{dT_{atm}}{dz} \tag{1.21}$$

This means the change in ambient temperature after moving from altitude $z_0$ to $z_1$ is given by:

$$T_{z1} = T_{z_0} + \Gamma_{ambient} dz \tag{1.22}$$

An atmosphere is considered convectively stable if this condition holds:

$$\Gamma_{adiabatic} > \Gamma_{ambient} \tag{1.23}$$

In the above equation, it is important to keep the signs of $\Gamma_{adiabatic}$ and $\Gamma_{ambient}$ and not take the modulus, so that this condition is correct.

If the condition of equation 1.23 holds, any vertically displaced parcel of gas increases or decreases in temperature more than its surroundings do. To clarify, if the parcel ascends it will expand, cool (more so than atmosphere) and reach an altitude where the environment is warmer. Thus, it will start to descend. Conversely if the parcel descends, it will contract and heat (more so than atmosphere) and reach an altitude where the environment is cooler. Thus, it will start to ascend. Therefore, there is always a restoring force opposing the parcel motion. This restoring force is strongest in the thermosphere due to its large positive temperature gradient. This means the atmosphere is not just decreasing in temperature at a smaller rate than the parcel temperature (the minimum requirements for stability) but it is actively increasing. This gives rise to a very strong restoring force and a particularly convectively stable atmosphere. Hence a lot of energy is required to lift the thermosphere vertically.

### 1.2.2 Composition

Figure 1.7 shows the concentration altitude profiles of the neutral and ionic species, clearly displaying the density of the neutrals as orders of magnitude larger than that of the charged particles. In the absence of vertical wind the vertical density variation is controlled by diffusive separation. The dominant ion species in the upper atmosphere are $O^+$, $NO^+$ and $O_2^+$ whereas the major neutral species in the thermosphere are atomic oxygen ($O$) and molecular nitrogen ($N_2$) with molecular oxygen ($O_2$) being more important at E region heights. At exospheric heights (>500km) the lightest elements helium and hydrogen become the dominant species while the density of atomic oxygen decreases. Generally, the

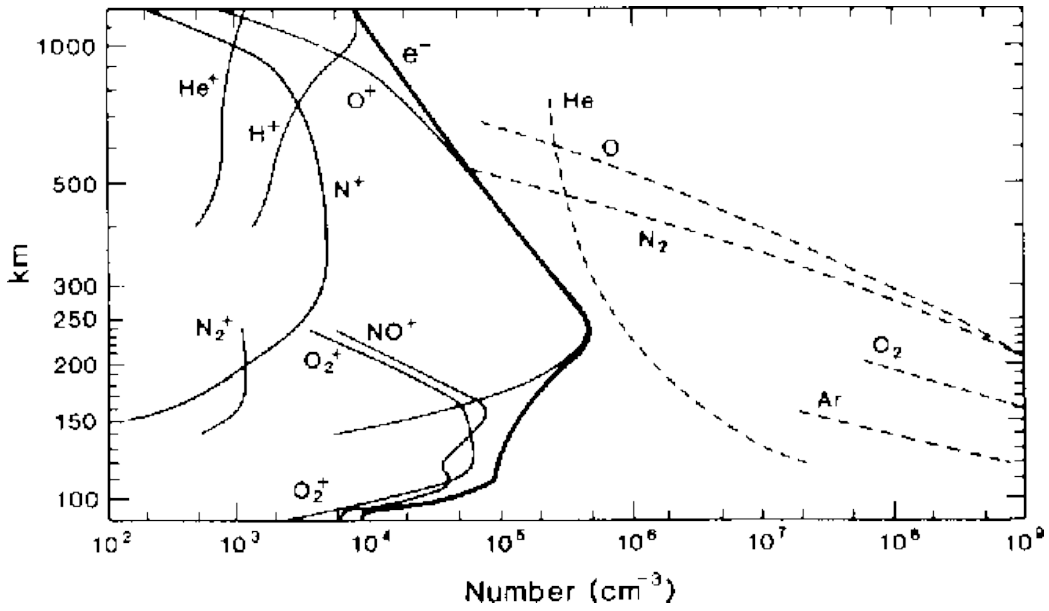


Figure 1.7: Density profiles of major atmospheric constituents from 100-1000km altitude (Johnson, 1969; Luhmann, 1995)

lighter constituents ($O$) fall off more slowly than the molecular constituents ($N_2$ and $O_2$) due to diffusive separation, e.g., the profile of $O$ is far steeper than that of $N_2$. However chemical processes can occur to intervene in this trend which will be highlighted in this section. A predominant lower thermospheric process and a principal production mechanism of atomic oxygen is the photodissociation of $O_2$ by solar UV:

$$O_2 + h\nu \rightarrow O + O \tag{1.24}$$

There are other production mechanisms of atomic oxygen (see section 1.5), which are present both day and night, however these arise from a coupling with the ionosphere. The main source of atomic oxygen at night-time for the E and F1 regions is via dissociative recombination of $O_2^+$ by thermal electrons:

$$O_2^+ + e^- \rightarrow O + O \tag{1.25}$$

The two chemical reactions above are responsible for the concentration of atomic oxygen exceeding its molecular counterpart above ∼110km altitude at a rate that cannot be explained purely by diffusive separation. In the absence of these two reactions, the drop-off of $O_2$ with height would be more similar to $N_2$ as their molecular weights only differ

by four atomic mass units and also as nitrogen is unlikely to undergo similar chemical reactions. Nitrogen has a small dissociation cross-section therefore is predominantly found in its molecular state and it's drop off with altitude is purely a result of diffusive separation. Its molecular ion state ($N_2^+$), however, has a larger dissociation cross section (see next set of reactions).

The ionosphere has been shown to be produced by various ionisation reactions involving incident photons and photoelectrons as described in section 1.1.3. The predominant ionisation processes in the ionosphere, however, help to explain its composition and have not yet been mentioned. The greatest source of ionisation in the lower thermosphere and the reason for the peaks in $NO^+$ and $O_2^+$ in the E and F1 region is due to absorption of solar EUV and X-rays which produces $NO^+$, $O_2^+$ and $O^+$. Since this flux is directly related to solar cycle variation, the relative proportions of these ions is depleted at solar minimum. The main night-time E and F1 region reactions are dissociative recombination with electrons:

$$e^- + O_2^+ \rightarrow O + O \tag{1.26}$$

$$e^- + N_2^+ \rightarrow N + N \tag{1.27}$$

$$e^- + NO^+ \rightarrow N + O \tag{1.28}$$

Additional key processes for the F1 region which contribute to atomic oxygen are charge exchange processes:

$$O^+ + O_2 \rightarrow O + O_2^+ \tag{1.29}$$

$$O_2 + O^+ + e^- \rightarrow O + O + O \tag{1.30}$$

The key F2 region process, almost completely responsible for the night-time ionosphere, and another source of atomic oxygen, is a two part dissociative recombination process between oxygen and nitrogen:

$$O^+ + N_2 \rightarrow NO^+ + N \tag{1.31}$$

$$e^- + NO^+ \rightarrow N + O \tag{1.32}$$

There is a plentiful supply of $O^+$ being a major species in the F2 region. The recombination rate, i.e., ionospheric loss, depends on the concentration of $N_2$ while the ionospheric production rate depends on the concentration of O, therefore the $\frac{O}{N_2}$ ratio is a good indicator of ionospheric density. As the oxygen drops off more slowly than nitrogen in the F region, the production rate drops off more slowly than the loss rate and so there is an increase in electron density with height (until the maximum is reached at ∼300km altitude) shown in Figure 1.7.

The more traditional explanation of the ionospheric density altitude profile is through the Chapman function. The solar X-ray and UV ionising radiation (which form the ionosphere) exponentially fall off as it passes through an absorbing medium, such as the Earth's thermosphere. In an atmosphere in graviational equilibrium such as the Earth's, there is an exponential decrease in density with altitude. Both the density and radiation intensity are proportional to ionisation. The exponential fall-off in density and exponential increase in ionising radiation with altitude, together form a layer of ionisation in the atmosphere defined by a function termed the Chapman function (Cummack, 1961), i.e., the ionosphere.

### 1.2.3 Circulation

**Ionosphere**

The ionospheric circulation at F region heights is primarily controlled by the magnetospheric dynamo imposing $\vec{E} \times \vec{B}$ drift upon the plasma. Due to increased neutral density in the lower atmosphere, the ions are increasingly controlled by the neutrals, via ion drag, separating from the electrons and forming currents. From F to D region altitudes, the ions are increasingly rotationally accelerated: in the near-collisionless F region, the ions are travelling at $0^o$ to the $\vec{E} \times \vec{B}$ (i.e., in the same direction); at some altitude in the E region the ions are travelling $45^o$ to the $\vec{E} \times \vec{B}$; and in the D region, the ions are travelling $90^o$ to the $\vec{E} \times \vec{B}$ in the direction of the electric field (Kosch et al, 2011). The ionosphere's circulation is also affected by the Coriolis force and the solar pressure gradient, however, these are minor influences compared to the electrodynamic ones.

The Coriolis force (for the ionosphere) is particularly dwarfed by the electrodynamic forces, e.g., the Lorentz force. This can be explained by the charge-to-mass ratio ($\frac{q}{m}$). If we take the ratio of the Coriolis acceleration (of the order $2\omega v$) to the Lorentz acceleration

(of the order $\frac{q}{m}vB$) this gives:

$$\frac{Coriolis}{Lorentz} \sim \frac{m}{q} \times 2 \times \frac{\omega}{B} \tag{1.33}$$

where $\omega$ is the angular speed of the Earth's rotation, B is the magnitude of the Earth's magnetic field and $v$ is the speed of a particle in a corotating reference frame. Considering that $2 \times \frac{\vec{\omega}}{\vec{B}}$ calculates to be close to unity and $\frac{m}{q}$ for an $O^+$ ion is $\sim 10^{-7}\text{kgC}^{-1}$, this shows that the Coriolis acceleration for a charged particle is a negligible fraction of the acceleration due to a magnetic or electric field.

**Thermosphere**

The global thermospheric circulation is driven by a mixture of driving forces. It is primarily controlled by the solar pressure gradient and the Coriolis force. However, at high latitudes, ion drag and momentum advection become more dominant. If the thermosphere is treated as a fluid subject to hydrostatic equilibrium, and assuming that there are sufficient collisions with respect to the time-scales of interest, a reasonably accurate equation describing the neutral velocity $\vec{U}$ is a modified version of the Navier Stokes equation which follows:

$$\frac{D\vec{U}}{Dt} = -\frac{1}{\rho}\vec{\nabla}P + \vec{g} - 2\vec{\omega} \times \vec{U} - \frac{\mu}{\rho}\vec{\nabla}^2\vec{U} - \nu_{ni}(\vec{U} - \vec{V}) \tag{1.34}$$

where $\vec{U}$ is the three dimensional wind velocity, t is time, $\rho$ is the density, $\vec{g}$ is the acceleration due to gravity, $\omega$ is the angular velocity, $\mu$ is the viscosity, $\nu_{in}$ is the ion-neutral collision frequency and $\vec{V}$ is the three dimensional ion velocity.

Each term in this momentum equation describes a different thermospheric driving force as follows:

- The first term in the equation is the material derivative of the neutral wind and is $\frac{D\vec{U}}{Dt} = \frac{\partial}{\partial t}\vec{U} + (\vec{U}.\vec{\nabla})\vec{U}$ and sums up the overall acceleration of the fluid with $\frac{\partial}{\partial t}\vec{U}$ being the time variation of $\vec{U}$ at a fixed point and $(\vec{U}.\vec{\nabla})\vec{U}$ being the acceleration of the fluid due to spatial gradients in $\vec{U}$. This term can also be called the momentum advection.

- The second term, $-\frac{1}{\rho}\vec{\nabla}P$, is the pressure gradient force which is created by various heat sources such as solar heating, Joule and particle heating and also waves and

tides dissipating or breaking (see section 1.3).

- The third term, $\vec{g}$, is the acceleration due to gravity which only has an impact on the vertical motion.

- The fourth term, $-2\vec{\omega}\times\vec{U}$, is the Coriolis force where $\vec{\omega}$ is the angular velocity of the rotating Earth. This is the apparent deflection in the movement of the neutral gas when the motion is described relative to the rotating reference frame of the Earth, i.e., a non-inertial reference frame. In the northern hemisphere the deflection acts to the right of the moving gas, and in the southern hemisphere the deflection acts to the left of the moving gas. It is strongest at the pole and weakest at the equator.

- The fifth term, $-\frac{\mu}{\rho}\vec{\nabla}^2\vec{U}$, is the viscosity term of the thermosphere which is a slow forcing term and is considered a minor driving force. This term is dominated by its vertical component as the vertical wind shears are so much larger than the horizontal.

- The sixth term, $-\nu_{ni}(\vec{U}-\vec{V})$, is the ion drag for an ion-neutral collision frequency $\nu_{in}$. There is a cut-off point in the thermosphere below $\sim$110 km where the neutral density becomes large enough so that the neutrals start to control the ion motion, rather than the ions controlling the neutrals, which occurs higher up in the atmosphere (e.g., Kosch et al, 2011).

Even though this momentum equation is three-dimensional, it can be adapted to describe the horizontal motion of the thermosphere. For the vertical component of this equation, the only significant terms are the pressure gradient force, the acceleration due to gravity and the viscosity term. Without the viscosity term and considering instances when there are no temporal gradients in velocity, the vertical momentum equation is equivalent to the hydrostatic equilibrium relation (see equation 1.15).

## 1.3 Energetics

The energetics of the thermosphere is a complicated aspect subject to many heat sources. A very simplified picture would describe the thermospheric temperature profile as a result of a balance between the absorption of solar radiation and the downward molecular conduction of heat (Killeen, 1987). In reality it is a far more complicated energy balance than this, particularly at high latitudes. The magnetospheric dynamo is a controlling feature.

The electric fields mapped down from the magnetosphere drive currents, causing energy to be collisionally transferred between the charged ionosphere and neutral atmosphere. Energy is also deposited via energetic particles being accelerated along magnetic field lines and subsequently decelerating or colliding with the neutral atmosphere, i.e., particle precipitation. The total power transferred from the magnetosphere into the ionosphere and thermosphere is typically hundreds of Gigawatts (Knipp et al, 2004).

Knipp et al (2004) use three empirical models to estimate the relative contributions from solar EUV, Joule heating and particle heating to the global energy budget of the upper atmosphere for solar minimum and maximum, and investigate proportions of each contribution by considering the most energetic heating events. Their study finds that solar EUV is the largest contributor, accounting for 80% of the budget providing ∼464GW, with Joule heating being the next most important providing ∼95GW and then particle heating providing ∼36GW. Joule and particle heating together contribute 17% of the global energy budget. In the top 10% and 1% largest heating events, this rises to 20% and 25%. Furthermore, if only considering the 15 largest heating events, these two geomagnetic heat sources combined contribute over 50%, exceeding the contribution from solar radiation. Therefore Joule and particle heating must not be underestimated. These different methods of heating are expanded on in the following subsections.

### 1.3.1 Solar UV, EUV and X-ray Absorption

Solar EUV and X-ray absorption is the most dominant and global-scale thermospheric heat source and is an important factor in the altitudinal temperature structure, particularly in the lower thermosphere. The chemical reactions that convert this energy into heat are strongly dependent on composition and thus on altitude (Torr et al, 1980). This can also be inferred from the many chemical reactions introduced in sections 1.1.3, 1.2.2 and 1.5.

The maximum thermospheric heating rate due to the absorption of UV and X-ray wavelengths (more specifically in the Schumann Rung continuum of 135-176nm), lies at 120km altitude. This is largely accountable to the dissociation of molecular oxygen. The excess energy within this reaction is given to the products as kinetic energy, raising their temperature. If the reaction is energetic enough sometimes one of the oxygen atoms is excited to a $^1D$ state (leading to 630nm emission). At these lower altitudes, the atom is quickly quenched, transferring the energy on to other species and raising their temperatures. This is the main heat source for the altitudes between 100-150km (Roble et al,

1987). In the upper thermosphere at F region heights, however, there is very little solar radiation absorption so this is not an important heat source here. For altitudes where it is an important heat source, it must be noted that the flux of solar radiation is dependent on the solar cycle. The solar radiation flux varies <1% with solar cycle; it is the UV and EUV, which are absorbed by atomic O, that have a strong solar cycle variation. Due to this variability, the thermospheric mean temperature, wind magnitude and diurnal difference in temperature is larger for solar maximum than minimum (Hernandez and Roble, 1995).

### 1.3.2 Joule heating

Joule heating is the frictional transfer of heat between ions and neutrals travelling at different velocities that occurs mostly at high latitudes. The difference in speed arises due to the balance of forces experienced by the ions and neutrals. To recap, the ions are accelerated by the electric fields mapped down from the magnetosphere which forces $\vec{E} \times \vec{B}$ drift in the F region. Lower down in the ionosphere, Pederson currents are driven and electric field energy is deposited via collisions with neutrals. The neutrals are more strongly influenced by the day-night pressure gradient and the Coriolis force, and through neutral-ion collisions, the ions gain a little of this momentum. Furthermore, the response times to changes in the IMF are different for the two atmospheric constituents: the ions are quick with a lag time of ∼2min (Heelis et al, 2002) and the neutrals are slower with a lag time in the range of 0.5-6.5hr (Kosch et al, 2001). Clearly, the ions and neutrals experience very different forcings and the ion drag is not strong enough to 'smooth out' velocity differences quickly, so there will always be some degree of Joule heating present in the upper atmosphere. The volume heating rate per unit mass due to collisions is given by (Brekke, 1997):

$$\frac{\partial q}{\partial t} = \frac{1}{2}\nu_{ni}(u_i - u_n)^2 \tag{1.35}$$

where $\nu_{ni}$ is the neutral-ion collision frequency, $u_i$ is the ion velocity and $u_n$ is the neutral velocity.

Maurice and Hanson (1982) describe the Joule heating rate with an additional term describing the thermal motion of the particles causing collisions and exchanging heat, as well as the collisional transfer from the bulk motion of the ion flow. Their version of the

Joule heating rate is given by:

$$\frac{\partial q}{\partial t} = \Psi_{ni} 3 k_B (T_i - T_n) + \phi_{ni} m_n (u_i - u_n)^2 \tag{1.36}$$

where $k_B$ is the Boltzman constant, $m_n$ is the neutral mass, $T_i$ is the ion temperature and $T_n$ is the neutral temperature. $\Psi_{ni}$ and $\phi_{ni}$ are functions which, to a good degree of approximation, are equal to 1 for all ion-neutral interactions (Maurice and Hanson, 1982).

The height-integrated Joule heating rate is often expressed in terms of the electric field and current density (Brekke, 1997):

$$\frac{\partial Q_J}{\partial t} = \int_z \vec{j}.\vec{E} dz \tag{1.37}$$

where $Q_J = \int_z q dz$, $\vec{E}$ is the electric field in the neutral rest frame and $\vec{j}$ is the current density vector. Above ∼100km altitude, the component of conductivity parallel to the magnetic field lines is so large that any charge imbalance is immediately neutralised and they act as equipotentials. As a result, the electric field consists only of a perpendicular (to magnetic field) component $E_\perp$. By using the component of conductivity parallel to $E_\perp$, i.e., the Pedersen conductivity, $\sigma_P$, and including the neutral wind term ($\vec{u_n}$), the height-integrated Joule heating rate can be written:

$$\frac{\partial Q_J}{\partial t} = \int_z \sigma_P(z) (\vec{E} + \vec{u_n} \times \vec{B})_\perp^2 dz \tag{1.38}$$

Due to the difficulties in measuring the neutral wind profile and the underestimated importance of it as a dynamo, a common simplification to equation 1.38 is:

$$\frac{\partial Q_J}{\partial t} = \int_z \sigma_P(z) \vec{E}_\perp^2 dz \tag{1.39}$$

where the neutral wind has been neglected.

The Joule heating rate peaks from 100-130km altitude near the peak of ionospheric current density, however, the maximum neutral gas temperature lies at ∼300km altitude. This is due to the lower neutral density at ∼300km altitude, which means the smaller Joule heating rate here (further from the peak of current density) is enough to create a larger energy per unit particle. At a lower altitude with a larger heating rate (closer to peak of current density) and a higher density there will be less energy per unit particle as

there will be so many more particles to distribute the energy between. Sometimes Joule heating will provide heat to the neutrals, however most of the time it is the ions which rise in temperature due to their smaller relative heat capacity.

Maurice and Hanson (1982) use Atmospheric Explorer C data-set to find a correlation between electric field enhancements and ion temperature when the DC electric field is above 20mV/m. They show a balance between heat exchange with neutrals and frictional heating in the ions. Kurihara et al (2008) find evidence of Joule heating above 120km altitude where the ion temperature increased by several hundred K in response to a large increase in electric field and an enhancement of the conductivity by hard particle precipitation (>keV).

Many studies miscalculate the Joule heating rate by excluding the neutral wind dynamo and also do not account for electric field variability. Thayer et al (1998) find the contribution of the neutral wind dynamo to vary with altitude. At some heights positive contributions are found, with the Joule heating rate being 400% larger than the wind-free calculation, and at other heights, negative contributions are found with the Joule heating rate being 40% of the wind-free calculation. They also find another phenomenon: when including the neutral wind dynamo the altitude region of Joule heating is confined. Aruliah et al (2005) use concurrent tristatic FPI and tristatic radar measurements and find a 29% contribution from the neutral wind dynamo at ∼240km altitude. They also show a 320% increase in the Joule heating calculation when using a 1-minute-average compared with a 15-minute-average of the wind. Condrescu et al (1995) argue that by including small-scale electric field variability, the Joule heating rate may increase by 33%. Rodger et al (2001) use EISCAT data to find a Joule heating rate underestimation of 20% when using the 1-hour resolution data compared with the 6-minute resolution. Furthermore, Deng and Ridley (2007) run the GITM model for two different grid sizes: $5^o$ and $1.25^o$ and find the higher resolution grid size to increase the neutral heating rate by 20% at 200km altitude. Considering mesoscale resolution, rapid and localised thermospheric responses to ionospheric changes can occur (Aruliah and Griffin, 2001) and should not be overlooked. This has a significant effect on ion-neutral momentum transfer and hence on the different speeds of the charged and neutral atmosphere, as well as the Joule heating and energy budget.

Joule heating can be described in two different ways: ion-frictional heating and Ohmic Joule heating. They both describe the same phenomenon, however, they are quantified

using different parameters: the latter uses forces and flows within the equations, and the former uses electric fields and currents (Strangeways, 2012). The Ohmic Joule heating form has been shown in equations 1.37-1.39 and is more frequently used in upper atmospheric calculations (compared to ion-frictional heating) due to the availability of electric field altitude profiles from incoherent scatter radar. It is particularly used at E-region altitudes where the currents are largest. Ion-frictional heating has been shown in equations 1.35-1.36 and describe Joule heating as a mechanical coupling between the ions and neutrals. It is less popular than the Ohmic calculation due to the difficulty in measuring thermospheric winds at all altitudes. It is favoured by FPI scientists who have access to FPI thermospheric wind data. Strangeways (2012) argues that the mechanical ion-frictional heating view of Joule heating is a more appropriate defintion, as, using the electric field as the starting point in describing the effect neglects causality, since the electric field is a consequence of the flow.

### 1.3.3 Auroral Particle heating

Particle precipitation is the entry of magnetospheric particles into the upper atmosphere along magnetic field lines. It is mostly confined to the region of the auroral latitudes where there is direct access for the energetic particles into the ionosphere. This is due to the auroral latitudes being a site of plasma flow shear across the open-closed field line boundary, hence from reconnection events, energetic magnetospheric particles can get accelerated to this region. Most of the energy is deposited between 100-200km altitude by collisions with neutrals. Particle precipitation is a more variable and localised phenomenon than Joule heating. It is often a likely explanation for sudden composition and temperature changes and is known to enhance the E region plasma density by an order of magnitude or more (Turunen et al, 2009). It is interlinked with Joule heating since, by ionising the neutral gas, it will also enhance the Pedersen conductivity. One must bear this in mind as they are nearly always treated as two separate phenomena.

## 1.4 Dynamics

### 1.4.1 Horizontal wind

The horizontal winds in the thermosphere are predominantly driven by the solar pressure gradient which can also be referred to as the diurnal tide. The pressure gradient force is

directed away from the point of maximum heating, centred around 14:00UT (or 17MLT for Svalbard). At high latitudes, this creates a flow over the pole to the nightside. At low altitudes the Coriolis force is more dominant so will force winds to flow along isobars, in a more geostrophic fashion (pressure gradient balanced by Coriolis force). Whereas at thermospheric heights, this effect is not so important and the winds flow perpendicular to isobars. E region winds maximise at roughly 100-110km altitude and can reach over 100ms$^{-1}$ whereas F region winds are larger, regularly exceeding a few hundred ms$^{-1}$.

At high latitudes ion-drag can play a significant role as charged particles drift under the effect of the magnetospheric convection electric field and of the geomagnetic field. Neutrals gain momentum from collisions with drifting ions and so in theory can start to follow the ions $\vec{E}\times\vec{B}$ drift. Aruliah and Griffin (2001) demonstrate that for low geomagnetic activity, $K_P$ <2 (the $K_P$ index is a measure of geomagnetic activity, defined in section 1.7.4), the solar heating is the dominant driver for the neutral wind whereas for moderately active geomagnetic activity, 2<$K_P$ <5, ion-neutral momentum transfer can play a significant role in driving the neutral wind and creating variability. Due to its large inertia, the neutral thermosphere can help to maintain any previous $\vec{E}\times\vec{B}$ pattern even after the cessation of strong magnetospheric driving. This has been termed the flywheel effect (Lyons, 1985).

The neutral wind is understood to be influenced by the IMF configuration and the resultant plasma drift. With a southward IMF Bz component and a strong twin convection cell ionospheric circulation, the cross-polar cap jet can reinforce the neutral solar pressure gradient flow over the pole. However at auroral latitudes, the ion drift turns antisunward and this will oppose the pressure gradient and reduce the magnitude of the winds here, possibly even rotating them. In another situation where the IMF Bz is northward, which is known to set-up a multi-cellular convection pattern, this can act to reduce the magnitudes of the (usually strong) polar antisunward wind. Under very strong northward IMF this polar flow can be reversed and flow in the sunward direction (Niciejewski et al, 1994).

### 1.4.2 Vertical Wind

Vertical winds play an important role in redistributing heat and energy as well as affecting the chemistry and general dynamics of the upper atmosphere. By diffusive separation the chemical composition is vertically stratified so a vertical motion can act to transport a particular species to new altitudes upsetting this chemical equilibrium. This alters the production and loss rates of different ionospheric species at different altitudes which can

result in a variation in the height of the ionospheric layers. Vertical winds are often accompanied by horizontally divergent flows due to conservation of mass, so can often act as the drivers of (or be driven by) horizontal winds.

Generally global-scale vertical wind motions are assumed to not exceed a few $ms^{-1}$ and are predominantly driven by solar heating of the atmosphere resulting in a diurnal 'breathing' of the whole atmosphere (Smith, 1998), i.e., a movement of gas along geopotential heights from the sun-side to the night-side of the atmosphere. At high latitudes more localised and energetic geomagnetic forcings such as particle precipitation and Joule heating become important which theoretically are known to violate hydrostatic equilibrium. These can also manifest as diurnal variations in vertical winds as the observing sites make their daily transit underneath geomagnetic heat sources. As geomagnetic activity increases so does the amplitude of the diurnal variation, particularly so for the downward wind (Smith and Hernandez, 1994; Conde and Dyson, 1994). For example, Conde and Dyson (1994) perform a statistical analysis on the F region vertical motions above Mawson Antarctica to find average quiet-time magnitudes of -2.6-3$ms^{-1}$ (minus indicates downward wind) which increases to -6.8-4.1$ms^{-1}$ at times of moderate geomagnetic activity. Within their statistical sample they also find many instances of winds in excess of 25$ms^{-1}$; however these are more often downward. Larson and Meriwether (2012) pool together various mid-latitude thermospheric vertical wind observations dating back to the 1960s taken either in situ, by rockets, satellites or ground-based FPIs. All these instruments have their characteristic advantages and limitations, compromising between time coverage, altitude range and data reliability, however they all show remarkably consistent results: thermospheric vertical winds with velocities of 10$ms^{-1}$ or more, typically lasting an hour or more in one direction, either upwards or downwards, for mid-latitude sites (20-50$^o$).

Aruliah et al (1994) carry out a statistical analysis using eight seasons of data from the Kiruna FPI which is situated at auroral latitudes. They confirm a dependence of thermospheric vertical wind on geomagnetic activity, solar cycle, UT and season. It is found that geomagnetically active nights show a more sustained period of vertical wind, sometimes up to 6 hours, with an average amplitude of 10$ms^{-1}$. For the quiet nights, on the other hand, the vertical wind oscillations are shorter and have smaller amplitudes, even though they can reach 10$ms^{-1}$. The fact that the average amplitude for an active night is 10$ms^{-1}$ indicates that winds considerably larger than this must frequently occur. At polar

latitudes the diurnal vertical wind pattern is found to be of a larger magnitude than can be explained by hydrostatic atmospheric 'breathing'. Rees et al (1984a) report a diurnal recurrence of a vertical wind pattern seen in the F region polar cap winds at Svalbard over seven days in 1980: a downward wind on the dayside from 06:00UT-18:00UT and an upward one on the nightside from 20:00UT-05:00UT, with average magnitudes 30ms$^{-1}$ and 50ms$^{-1}$ respectively, considerably larger than usually reported. A similar diurnal pattern is reported by Smith and Hernandez (1994) using upper thermospheric data from the South Pole however with amplitude 40ms$^{-1}$ in the F region and 10ms$^{-1}$ in the E region. This behaviour is expected as this South Pole site and Svalbard share very similar horizontal wind systems due to their similar locations with respect to the auroral oval. Together with the aforementioned studies, these findings seem to suggest that, generally, the magnitude of wind speed is larger the closer to the pole (and thus further from the auroral oval) the observer is situated, which is counterintuitive. In chapters 4-5 there is an in depth investigation of this phenomenon.

Many large-amplitude vertical motions are reported at high-latitude sites. However, these tend to be more localised and short-lived with spatial scale sizes of tens to hundreds of km and temporal scales of tens of minutes to a few hours. Kurihara et al (2009) observe localised upward winds of 40ms$^{-1}$ in the E region associated with temperature enhancements. Smith and Hernandez (1995) observe extreme vertical winds of 150ms$^{-1}$ on the most active day in their statistical sample for their South Pole station. Conde and Dyson (1995) show many examples of winds in excess of 100ms$^{-1}$ in their statistical sample. Spencer et al (1982) observe winds ranging from 100-250ms$^{-1}$ at auroral oval latitudes.

Statistical analyses on vertical wind magnitudes have been reported using a frequency distribution histogram. Anderson et al (2012) compile a vertical wind frequency distribution for 19 days of bistatic 630nm FPI vertical wind observations at auroral latitudes and fit a Gaussian profile significantly with a mean of $\mu$ =-0.63($\pm$0.03)ms$^{-1}$ and a standard deviation of $\sigma$ =15.25($\pm$0.03)ms$^{-1}$. Conde and Smith (1994) use their 103-night statistical sample of red line spectra to create these same histograms for $K_P$ <3 and for $K_P$ >3 ($K_P$ definition in section 1.7.4) and find distributions roughly centred about zero, both with a downward skew. This asymmetry is more pronounced for the quiet-time histogram which, due to Mawson being expected to lie just equatorward of the auroral oval at quiet times, is consistent with the theory that, adjacent to the auroral oval, the air subsides (i.e., more

downward vertical wind).

**Scale Sizes**

Much work has been done on the horizontal scale sizes of vertical wind flow patterns, with distances ranging between 10s-100s of kilometres observed. However, the scale of the flow is still not a well understood quantity. There is a great inconsistency between all estimations from various studies.

Spencer et al (1982) use DE2 satellite data to report strong vertical winds between 300-500km altitude which extend horizontally over 400km. Crickmore (1993) use a single-station FPI at Halley Antarctica to estimate a horizontal scale size of 360km. Kosch et al (2000) cross-correlate vertical wind measurements from two stations in Norway, separated longitudinally by 45km, to find no correlation for both E and F regions. However, some correlation between the upper and lower thermosphere is observed, indicating that the vertical scale size can span ∼100km, whereas the horizontal scale size can sometimes be less than 45km. This study indicates the presence of small scale transient structure driving strong vertical winds.

Price et al (1995) use Poker Flat, Alaska FPI observations of 630nm and 557.7nm emissions from 21 nights to find simultaneous E and F region upwellings lasting between 12-25 minutes on the poleward side of the auroral oval. These are strong vertical winds with a peak velocity of 42ms$^{-1}$ at E region height and 138ms$^{-1}$ at F region height and they are accompanied by a temperature increase of 200K at the lower altitude. They estimate an upper bound for the vertical wind horizontal scale size of 800km.

Ishii et al (2001) find a poor correlation between observations of F region vertical winds at two stations separated by 300km in Norway, showing this to be an upper limit to the scale size. On the other hand, the E region observations showed some degree of correlation and a correlation was also found between the upper and lower thermosphere vertical winds at both stations. Anderson et al (2011) uses bistatic Doppler measurements with two FPIs in Antarctica to find correlated vertical winds at separations of ∼ 150-480km and strong upward winds poleward of aurora.

Pooling all this scale size research together shows that, often, but not always, a correlation between adjacent stations is found. It seems that the scale size is typically ∼200km - however, it can range from less than 50km to as large as 800km, with extreme upwellings likely to have a vertical scale size of >150km.

**Large Vertical Winds**

The large vertical winds observed by Spencer et al (1982), Crickmore et al (1991) and Innis et al (1996) are on the poleward edge of the auroral oval, where there were more often upward winds, whereas the downward wind was more commonly observed at the equatorward boundary. This phenomenon is investigated and displayed more explicitly by Greet et al (2002) in figure 1.8) which shows the average vertical wind vectors distributed around the auroral oval in MLAT-MLT (a coordinate system allowing a bird's-eye views of the magnetic north pole) maps. In figure 1.8 the size of the symbol represents the speed of the vertical wind which reveals that the larger magnitude winds are found on the poleward side of the oval compared to those within the oval. Innis and Conde (2002) perform a statistical analysis and find the highest vertical winds in the auroral oval at high altitudes and especially when the AE index (a measure of auroral electrojet activity, calculated using data from magnetometers across the auroral oval) was larger. However the magnitude in their study was independent of solar zenith angle.

Rees et al (1984a) suggest that downward winds usually occur adjacent to the auroral oval as the convective subsidence from the upward winds within the auroral oval itself. Auroral energy deposition enhances the neutral temperature directly, causing a sudden rise in local temperature which causes these vertical winds to be generated. Changes in vertical wind seem to arise instantaneously due to changes in the heating rate (Kurihara, 2009). Oyama et al (2008) derive neutral vertical wind speeds using EISCAT data and find magnitudes exceeding $30\text{ms}^{-1}$ accompanied by an auroral heating event. Guo and McEwen (2001) observe upward vertical wind during and following a sustained aurora above their polar cap site, suggesting that heating due to particle precipitation is sufficient to sustain upwards flow. Rees et al (1984a) identify a source of intense local heating with a period of 15-30mins in the magnetic midnight region driving rapid upward vertical motion accompanied by intense magnetic disturbances, evidence of the MIT system at work. Sometimes a counter-intuitive drop in temperature can accompany large vertical winds driven by particle precipitation. This is due to particularly hard precipitation causing intense aurora and forcing the emission towards a lower altitude (Ishii et al, 2001; Holmes et al, 2005).

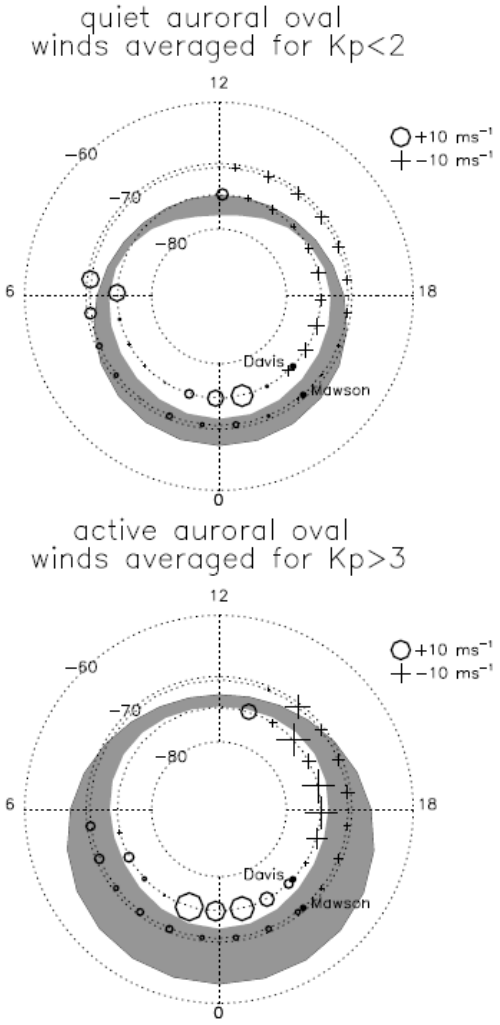


Figure 1.8: South pole distribution of positive and negative vertical winds with respect to the auroral oval at geomagnetically quiet (top) and active (bottom) conditions (Greet et al, 2002)

**Burnside Relation**

A relationship between vertical wind velocity and divergence of horizontal wind velocity must exist due to conservation of mass, and under special circumstances this relation is the Burnside relation:

$$\vec{w} = H(\vec{\nabla} \cdot \vec{u}) \tag{1.40}$$

where $\vec{w}$ is the vertical wind, $H$ is the pressure scale height and $\vec{u}$ is the horizontal wind vector. The full derivation can be found in chapter 6.

Historically, the Burnside relation has been of great interest to the FPI community due

to the difficulty in measuring the small magnitude vertical winds ($\sim$10ms$^{-1}$) compared to the larger horizontal winds ($\sim$100ms$^{-1}$); this offered a more precise alternative method of measurement. Burnside et al (1981) found the precision of the vertical wind estimations to be ten times greater using the Burnside relation rather than by direct measurement. Now with our modern more sensitive FPI detectors, this is no longer the case. However, the relation is still valued by the FPI community. If the relation is upheld in the thermosphere, it can be used to infer a vertical wind field across different zenith observing angles by a single instrument, revealing information regarding the horizontal scale size of the vertical wind; something that would be invaluable for the SCANDI instrument's wind analysis (see section 3.11).

For the Burnside relation to hold, the atmosphere at the height of the emission is constrained to be under stable conditions, i.e., small vertical winds, slowly varying pressure and with an isothermal atmosphere above. This is typically the situation in the thermosphere particularly at low-latitude sites where Burnside et al (1981) used this relationship successfully for the first time. At high-latitude sites, there can be deviations from these stable conditions due to energetic geomagnetic forcing. In the high-latitude observational studies so far (Guo and McEwen, 2003; Anderson, 2011; Crickmore, 1991), a strong positive correlation is often found between the vertical wind and horizontal divergence. However, the scale height is usually found to be 3-5 times too large. We note that in the work of Anderson (thesis, 2011), a reasonable range of scale heights is derived using data from their auroral site at Poker Flats.

The Burnside relation has great potential for inferring indirectly the vertical winds using horizontal wind measurements. If our FPI database displays a strong, well-defined relationship between the two wind components, either with a realistic scale height (complying with the Burnside relation), or without, this relation would be a very useful tool. This aspect is further investigated in chapter 6.

### 1.4.3 Small-Scale Structure

Due to its large inertia, a simplified view of the thermosphere is that it is a relatively smoothly varying (large spatial and temporal scales of $\sim$1000km and hours, respectively) medium which hosts the dynamic ionosphere. However, there is increasing evidence that small-scale spatial and temporal structure is a typical attribute of the thermosphere. Significant small-scale structure is often reported (e.g., Aruliah and Griffin, 2001; Conde and

Smith, 1998) and the thermosphere is found to be readily responsive to the ionospheric forcings. It is important to understand the full extent of the variability of the neutral atmosphere as the difference in the ion and neutral flow is a useful quantity in calculating crucial parameters concerning upper atmospheric energetics such as ion drag and Joule heating. Due to the limited spatial coverage of neutral wind measurements, 'smoothed-out' models with low spatial resolution, e.g., CMAT2 (Coupled Mesophere And Thermosphere, version 2), are used which give unrealistic model simulations. These modelled neutral winds (and ion velocities which are driven by the electric field) lack the observed small-scale fluctuations and so underestimate Joule heating (Qjh). This can be understood mathematically by considering the squared terms in equations 1.35 and 1.39. These result in larger Qjh estimations if an average of many Qjh values (high variability) is calculated rather than calculating just one average Qjh value over the same time period. Furthermore, by excluding the high variability of the electric fields from the Joule heating calculations, which causes unrealistically smooth ion velocity calculations, this acts to overestimate the momentum transfer. This is because the high variability of the ion velocity results in the ion velocity fluctuations to be randomly directed, rather than all in the same direction (as is seen more with a smoothed ion velocity signal). Thus, the net momentum transferred is less than if all the ions (driven by E) are moving in the same direction (which is what is assumed with steady state E-field models, e.g. CMAT2).

Neutral winds are variable across different altitudes and it is notoriously difficult to derive complete altitude profiles, so they are often assumed to equal zero - which is a major source of error. Thayer et al (1998) found that the Joule heating can be up to 400% different when including the neutral wind in the calculation.

## 1.5 Airglow, Aurora and Atomic Oxygen Emission

Airglow is the global emission of weak diffuse radiation by atmospheric species via photochemical reactions, e.g., radiative recombination and ionisation reactions. The terminology further splits into nightglow and dayglow to distinguish between emissions that occur in the day or the nightside, whereby at each of these regions different reactions assume different levels of importance. For dayglow, the emissions will be obscured by the solar radiation which is many orders of magnitude brighter and for nightglow in high-latitude regions there will be an additional source of radiation from auroral emissions.

Auroral emissions occur at high latitudes and are highly structured and more intense than airglow. Since their production mechanism is energetic electron impact, their peak emission height is very variable as it is highly dependent on the incident electron energy. Airglow and auroral emissions range in wavelength from the far infra-red to the far-UV and are produced by photochemical reactions which result in excited atmospheric species. As long as these species do not undergo collision with other species (quenching) then this excitation energy is lost via radiation, thus, forming airglow and auroral emissions. The two predominant auroral emissions are from atomic oxygen: the green line (557.7nm) emission from the $^1S_0$ - $^2D_2$ transition and the red line (630nm) emission from the $^1D_2$ - $^3P_{2,1,0}$ transition.

The green line emission peak has a variable height due to the different mechanisms of production but generally lies in the E region. A good average emission altitude is ∼110km. The red line, however, has a more stable emission height at ∼240km altitude due to the reactions shown later in this section. The green line has a lifetime of ∼0.91 seconds whereas the red line has a lifetime of ∼110 seconds. The altitudes of these two emissions conveniently probe regions that are too high for direct balloon measurements but are too low for a stable satellite orbit. Therefore, by measuring these emissions with FPIs, we acquire a unique insight into this region of the thermosphere. In this thesis we focus on the red line emission as it is has a more stable emission altitude and is situated in the F region which is the region of interest for this study.

The red line transition is in fact an amalgamation of three different subtransitions between the excited state $^1D_2$ to three possible ground states $^3P_2$, $^3P_1$ and$^3P_0$ which produce emissions at 639.2nm, 636.4nm and 630.0nm. The first subtransition is too weak to be observed (Whitten and Papoff, 1976) and the remaining two give a brightness ratio 630: 636.4 = 3: 1 (Rishbeth and Garriott, 1969). Therefore, it is reasonable to assume the transition is purely 630nm and neglect the other two subtransitions. The 630nm emission occurs via a forbidden transition in accordance with the electric dipole selection rules therefore can be termed a metastable state. This means it has a low probability of occurring and a long radiative lifetime of 110 seconds. This leaves the excited oxygen atom in danger of being collisionally deactivated or quenched. This is a commonplace phenomenon at lower altitude regions which explains why there is little red-line emission. There are several production methods of the $^1D_2$ state of atomic oxygen (Rees and Roble, 1986) which leads to 630nm emission.

At daytime the dominant mechanism is the photodissociation of $O_2$ by solar radiation in the Schumann-Runge continuum:

$$O_2 + h\nu \rightarrow O(^1D_2) + O(^3P) \tag{1.41}$$

At night in the absence of aurora, the dominant mechanism above ∼280km altitude is dissociative recombination of $O_2^+$ ions:

$$O_2^+ + e^- \rightarrow O(^1D_2) + O \tag{1.42}$$

An important reaction, particularly for times of strong magnetosphere-ionosphere coupling at auroral regions, is the energetic electron impact excitation:

$$O(^3P) + e^*(E > 1.96eV) \rightarrow O(^1D) + e^- \tag{1.43}$$

An overall minor production mechanism but a dominant one below 200km altitude is the production of nitric oxide:

$$N(^2D) + O^2 \rightarrow O(^1D_2) + NO \tag{1.44}$$

Another minor production mechanism is the cascade reaction which produces the green line. This is an important source at lower altitudes (Solomon et al, 1988):

$$O(^1S) \rightarrow O(^1D) + h\nu(\lambda = 557.7nm) \tag{1.45}$$

Two other minor production mechanisms are quenching of atomic nitrogen and impact dissociation of $O_2$ which follow:

$$N^* + O \rightarrow O(^1D) + N(^4S) \tag{1.46}$$

$$O_2 + e^* \rightarrow O(^1D) + O + e^- \tag{1.47}$$

## 1.6 The Cusp

The cusp region is the region of the upper atmosphere (and magnetosphere) where the entry of magnetosheath plasma is most direct. It is situated at $\sim 78^o$ latitude, $\sim$2 hours around magnetic noon extending $\sim 4^o$ latitudinally. The characteristic high flux of soft precipitating particles ($\sim$hundreds eV) is due to the dayside reconnection causing a 'dribbling in' of particles from the magnetosphere.

A localised signature of the cusp is a density bulge at 400km altitude which is a few hundred km wide, ranging from 1.04-2 times the surrounding density (Lühr et al, 2005; Sadler et al, 2012; Zhang et al, 2012; Kervalishvili and Lühr, 2013; Schlegal et al 2005). This is fed by an upwelling from the atmosphere underneath (Carlson, 2012). This relatively new finding showcases another example of mesoscale variability in the thermosphere. Modelling this feature has presented a challenge to modellers (Demars and Schunk, 2007; Schlegal et al, 2005) who are forced to input unrealistic quantities of energy in order to replicate the bulge.

The generation mechanism of the density bulge is not yet agreed upon. The two main possibilities are Ohmic Joule heating in the E region and ion-frictional drag heating in the F region, particularly in the upper-F region where the atmosphere is more tenuous and easier to lift (Carlson, 2012). These possibilities are explored further in chapter 7 on the cusp.

## 1.7 Instruments

Ground based optical instruments, e.g, FPIs, are the most frequently used for measuring thermospheric winds due to their extended time coverage. Their drawbacks are that they are: i) limited to narrow altitude and latitude ranges, ii) operation hours are limited by cloud cover and darkness, and iii) atmospheric extinction effects require extra data calibrations (see section 3.9.1). The in situ satellites and rockets do not have these problems; they provide good vertical coverage and can cover a range of latitudes. However, satellites are limited to higher altitudes and burn up quickly when approaching 200km altitude. Rockets with chemical releases are the only technique to bridge the gap from 150-225km altitude. The major drawback to these in situ measurement techniques is the poor temporal coverage.

### 1.7.1 UCL's FPI network

UCL operates a network of FPIs in northern Scandinavia which measure three neutral parameters of the thermosphere: emission intensity, neutral wind velocity and neutral temperature. FPIs work by observing a single atomic transition from the sky field of view and passing it through a narrowband filter, reflecting the light between two etalon plates so as to obtain concentric interference fringes, which the instrument's software translates into spectra. The line-of-sight (LOS) wind speed, neutral temperature and intensity are inferred from the Doppler shift, Doppler broadening and the integrated area of the spectral profile respectively.

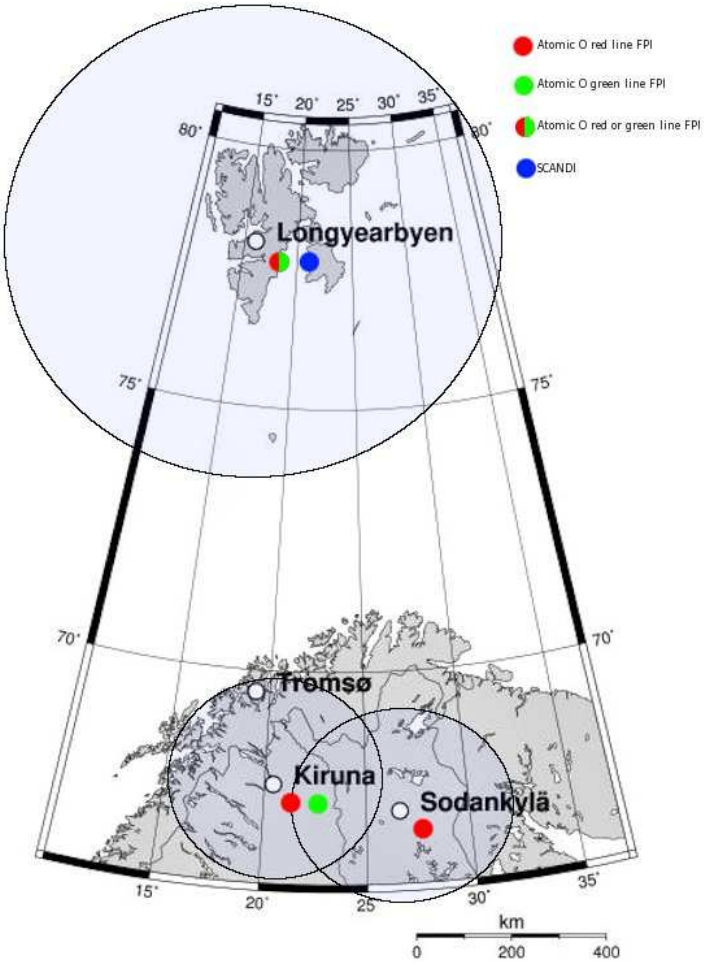


Figure 1.9: UCL's FPI network showing their FOVs. There is a red & green line FPI in Kiruna Esrange Optical Site (KEOPS), Sweden (67.8N,20.4E), red line FPI in Sodankylä, Finland (67.4N,26.6E), either red or green FPI and SCANDI in the polar cap at Longyearbyen in Svalbard (78.2N,15.6E). This image belongs to EISCAT (find at http://kaira.sgo.fi/2011/03/eiscat-svalbard-radar-part-1.html) however the FPI FOVs and coloured dots have been added.

UCL's FPIs measure either the 630nm atomic $O^1S$ red line emission or the 557.7nm atomic $O^1D$ green line emission, with peak emission intensities at 240km and 120km altitude respectively. The derived data will be an average in altitude, weighted according to the optical emission intensity profile along the observers LOS. Figure 1.9 shows the FPI network. There are five instruments marked by coloured dots (indicative of the colour emission measured) and their fields of view are marked by a dimmed circle. There are three FPIs located in the auroral oval: a red and a green line FPI in Kiruna Esrange Optical Site (KEOPS), Sweden (67.93N, 21.07E) and a red line FPI in Sodankylä, Finland (67.37N, 26.63E). These all have a 1° FOV (which can be directed towards various look directions), and they have a fixed observation zenith angle of 45°. This means the distance between two look directions positioned opposite each other (separated $180^o$ azimuthally) across the FPI's field-of-regard (i.e., the circular region of the sky which encompasses all the FPI look directions), is ∼500km across for observing the red line and ∼240km across for observing the green line. In the polar cap, at Longyearbyen, in Svalbard (78.15N, 16.04E), there is a standard FPI (narrow field) and a wide-field FPI (SCANDI). Both can measure either the red or green atomic O emission. The Svalbard FPI also has a 1° FOV which can be directed towards various look directions. The look directions are set at a fixed 60° zenith angle (giving a field-of-regard ∼870km across for observing the red line and ∼415km for observing the green line). SCANDI (blue dot) has a 150° field-of-view with a maximum observation zenith angle of 67°. This maps out to a circular region of the sky ∼1200km across for observing the red line (∼10° latitude across) and ∼565km across for observing the green line (∼5° latitude across).

SCANDI operates using the same principles as a standard FPI however it allows simultaneous mesoscale measurements over a FOV with a horizontal diameter of just over 1000km at an altitude of 240km with a time resolution of 8 minutes. The image of the sky is scanned by varying the gap of the etalon. The operation of SCANDI is described in detail in section 3.2.

The UCL FPI network is located conveniently close to other ground-based ionosphere/thermosphere measuring instruments. The Svalbard instruments' FOVs overlap with the European Incoherent Scatter Scientific Association (EISCAT) Svalbard Radar (ESR), one of the Super Dual Auroral Radar Network (SuperDARN) radars (Chisham et al, 2007), and the University Centre in Svalbard (UNIS) Meridian Scanning Photometer (MSP). The Kiruna and Sodankylä FPIs have two common observing volumes, named

the tristatic A (for an experiment when there was a third FPI at Skibotn, run by Lancaster University) and the bistatic B look direction (the northern intersection of the FOVs of the Kiruna FPI and the Sodankyla FPI). The mainland EISCAT radar system can be organised to observe bistatic (or tristatic) look direction observations at this same common volume. These experiments yield concurrent neutral and charged atmosphere measurements.

### 1.7.2 Statistical Analysis and Data Availability

UCL's long historical FPI data base of thermospheric measurements (from operating instruments at many different sites over many years) allows statistical analyses to be carried out. These are useful in determining the typical behaviour of the thermosphere rather than following a (potentially one-off) thermospheric event as a case study. The auroral emissions are only detectable at night time when there is no background solar emission. This means UCL's instruments only take measurements throughout winter from mid-September to mid-April. There are many reasons for data not being collected on a night within the observing season such as: instrument failure, computer crashes and power cuts.

Cloudy nights may also reduce the useable database significantly. Under cloudy conditions, the airglow emissions are scattered in every direction so directional information for each photon is lost. This creates spurious Doppler shifts. Therefore, data from cloudy nights is considered unreliable. Cloud cover records are produced, involving searching through 1 minute resolution All Sky Camera (ASC) images and observing the sky conditions directly. Before the ASCs were installed (January 2002 for KEOPS and November 2007 for Svalbard) cloud cover was estimated by inspecting the wind and intensity plots from all the look directions. With experience, the presence of cloud scatter can be estimated, during geomagnetically quiet conditions, by looking for near identical intensities in all directions and small, raggedy wind variations in the horizontal look directions caused by the cancelling out of red and blue Doppler shifts. When there is a large intensity gradient, due to the presence of the auroral oval, the same Doppler shifts are seen in all directions, including the zenith. Either way, the average observing conditions for each hour are translated into a code in a cloud record file. I created these files for many months of each season (2002-2014). These files can be read into data analysis programmes and used to select just the clear weather data for statistical analyses.

### 1.7.3 Supplementary Instruments

#### All Sky Cameras

All UCL's FPIs are co-located with $180^o$ fish-eye lens All Sky Cameras (ASCs) which take broad-band images of the entire sky every minute in order to log the weather conditions in the FPI field of view and to image any auroral activity. The Electron Multiplying Charge Coupled Device (EMCCD) chips within the cameras are highly sensitive in order to detect the faintest aurora. Figure 1.10 shows a sample ASC image taken whilst a substorm was in progress. The first ASC was installed in January 2002 and has been operating routinely since. Additional ASC instruments were installed in Sodankylä and Svalbard in November 2007, as well as one at Kiruna in February 2015.

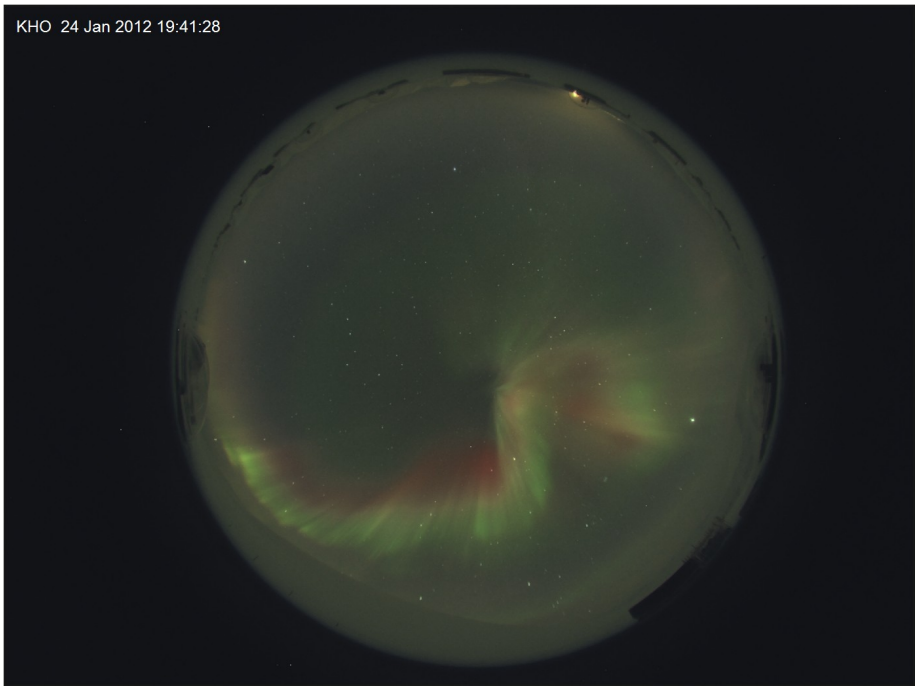


Figure 1.10: Svalbard ASC image taken at 19:41UT on 24th January 2014.

#### SuperDARN

Super Dual Auroral Radar Network (SuperDARN) is an international radar network compromised of 35 high frequency (3-30MHz) coherent radars (Chisham et al, 2007). They are located at high latitudes in the southern and the northern hemisphere. Spherical harmonics are fitted to the data in order to create maps of high latitude plasma convection in the F region of the ionosphere (see figure 1.4) with the method shown in (Ruohoniemi

et al, 1998). A section of the map is covered by the Svalbard FPI and SCANDI field of view.

**EISCAT**

European Incoherent Scatter Scientific Association (EISCAT) operates three incoherent radar systems situated in Scandinavia. They provide ionospheric parameter data: plasma drift velocity, electron density and ion and electron temperature, from which, altitude profiles can be created. One is situated in Tromsø and the other two are collocated with UCL's FPIs in Kiruna and Svalbard.

**ACE**

ACE (Advanced Composition Explorer) is a NASA space-craft launched in 1997 with a mission to study the solar wind in the interplanetary medium. It operates close to the L1 Lagrange point between the Earth and the Sun lying ∼900,000 miles in front of the magnetopause. Onboard is the Magnetometer Instrument (MAG) which is constantly measuring the IMF $\vec{B}$ carried by the solar wind before it comes into contact with the magnetopause and interacts with the Earth. The separation of the magnetic measurements into Bx, By and Bz is used to infer the configuration of the high latitude electric field patterns.

**IMAGE**

IMAGE (International Monitor for Auroral Geomagnetic Effects) is a network of 31 magnetometer stations scattered across Scandinavia operated by 10 institutes from Norway, Sweden, Finland, Germany, Poland, Estonia and Russia spanning 58-79$^o$ latitude as can be seen in figure 1.11.

IMAGE monitors perturbations of the local magnetic field from its geomagnetically quiet state condition in three components: Bx (positive north), By (positive east) and Bz (positive vertically downwards). These are essential for the calculations of electric field, current density, ion drag, conductivity and Joule heating. They can be used to identify a substorm onset. During substorms the location of the electrojet can be determined by studying the northward component Bx and the vertical component Bz. Directly underneath the peak current of the electrojet, the magnetometers will measure a maximum Bx component with $\Delta B_x > 0$ for an eastward current and $\Delta B_x < 0$ for a westward current.

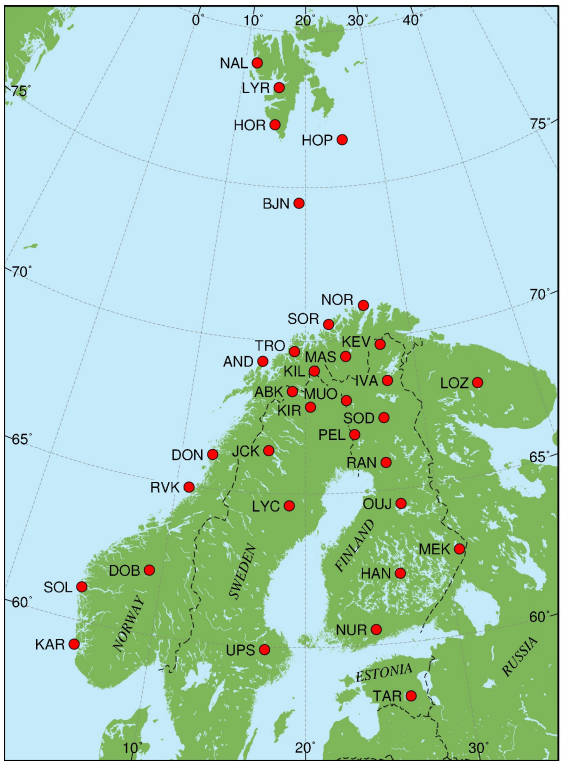


Figure 1.11: IMAGE magnetometer array in November 2014 courtesy of IMAGE official website (http://space.fmi.fi/image/beta/?page=maps).

The $\Delta B_z$ component will be zero. On either side of the electrojet current, the magnetometers will measure a minimum perturbation in Bx and a maximum perturbation in Bz.

#### Meridian Scanning Photometer

The Meridian Scanning Photometer (MSP) is a photometer built and operated by UNIS (The University Centre in Svalbard). It is combined with a rotating mirror so that it can scan along the magnetic meridian, hence the instrument's name. It is capable of measuring five different atomic emissions using five different narrow bandpass filters which are mounted on a tilted frame in front of a cooled photomultiplying detector. The instrument is designed to measure the absolute intensity of these emissions (in Rayleighs) as a function of elevation angle from north to south. It measures three atomic lines of oxygen, the two that UCL's FPIs measure: 557.7nm and 630.0nm, and also one at 844.6nm. In addition, the 427.8nm of molecular nitrogen and the 486.1nm of hydrogen emission lines are measured. The MSP is situated in the same building as UCL's FPIs in Svalbard, Kjell Henriksen Observatory, an ideal location for observing the day-side auroral cusp region.

The use of the MSP data (from the Kjell Henriksen Observatory (KHO) on Svalbard, Norway) within this thesis is courtesy of Prof Dag Lorentzen at UNIS.

### 1.7.4 Other Data and Measurements

#### Solar radio Flux F10.7

F10.7 is a measure of the solar radio flux at 10.7cm. It is used as a proxy for solar activity as it is found to be well correlated with sunspot number, visible and UV irradiance. This index has been measured since 1947 and is used as a measure of the diffuse, non-radiative heating of coronal plasma trapped by magnetic fields over active regions (Tapping, 1987).

#### $K_P$ and $A_P$ Indices

$K_P$ and $A_P$ indices are all measures of disturbance in the horizontal component of Earth's magnetic field. They are derived from the maximum disturbance in the horizontal geomagnetic field in a 3-hour interval. The $K_P$ indices range from 0-9 with 1 representing geomagnetically quiet conditions and >5 representing active conditions. $A_P$, is the same measure as $K_P$, however on a linear scale so different indices may be averaged in this scale and converted back to $K_P$. $A_P$ ranges from 0-400 with 1 representing calm conditions and >48 representing active geomagnetic conditions. Although these geomagnetic activity indices have been criticised recently as not being an adequate representation, they have the advantage of a very long database back to the early 1930s.

Following is a table showing the equivalent range of $A_P$ for a given $K_P$:

| $K_P$ | 0 | 0+ | 1- | 1o | 1+ | 2- | 2o | 2+ | 3- | 3o | 3+ | 4- | 4o | 4+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $A_P$ | 0 | 2 | 3 | 4 | 5 | 6 | 7 | 9 | 12 | 15 | 18 | 22 | 27 | 32 |
| $K_P$ | 5- | 5o | 5+ | 6- | 6o | 6+ | 7- | 7o | 7+ | 8- | 8o | 8+ | 9- | 9o |
| $A_P$ | 39 | 48 | 56 | 67 | 80 | 94 | 111 | 132 | 154 | 179 | 207 | 236 | 300 | 400 |

## 1.8 Models

### 1.8.1 CMAT2

CMAT2 (Coupled Mesophere And Thermosphere, version 2) is a general circulation model based on the assumption of hydrostatic equilibrium. It consists of a 3D grid over latitude, longitude and pressure level with 60 second time-step and runs by solving the equations of continuity, momentum and energy using a finite-difference scheme. It includes the

chemistry of major constituents, i.e., $N_2$, $O$, $O_2$ and $O_3$, and also has the option of minor constituent chemistry. It is coupled to the plasmasphere model that solves for $\vec{E} \times \vec{B}$ drift. The solar energy inputs are driven by radiative heating (F10.7 as a proxy), Joule heating (electric fields supplied by semi-empirical models - Foster or SuperDARN) and particle precipitation as energy deposition supplied by the $K_P$ which refers to the TIROS model of energy deposition (Fuller-Rowell and Evans, 1987).

### 1.8.2 MSIS

The Mass-Spectrometer-Incoherent-Scatter (MSIS) model is a semi-empirical global atmospheric model that extends from ground to space. The inputs required are time, date, altitude, latitude, longitude, local time, 81-day-averaged F10.7 values, daily F10.7 values and daily magnetic index. The model outputs various atmospheric parameters such as densities of $O$, $O_2$ and $N_2$ and the neutral temperature, density and exospheric temperature. It is known to be a fairly crude model yielding 'smoothed-out' parameters. However, it is a fairly good representation of ambient atmospheric conditions and hence is a useful reference tool to identify unusual behaviour in data.

This page was intentionally left blank

# Chapter 2

# The Fabry Perot Interferometer

## 2.1 Introduction

Fabry Perot interferometers (FPIs) are instruments used to filter and measure the wavelength of light. Developed by Fabry and Perot in 1901, they were first used to measure sky emissions in 1923 (Babcock, 1923) and now are widely used in spectroscopy, telecommunication and laser technology.

FPIs are highly suited to measuring sky emissions for two reasons: their high sensitivity and fine spectral resolution. High sensitivity is necessary in order to detect the weak airglow and auroral emissions. The red line is particularly weak, typically 10-50kR (Shemansky and Vallence Jones, 1968) and the green line is slightly more intense, typically ∼100kR (Vallance Jones, 1964). For context, a 1000kR aurora would have a brightness comparable to that of the light of the moon (Hargreaves, 1979). This low intensity is the reason for restricting measurements to night-time, as the signal is swamped by solar continuum during daylight. Secondly, high spectral resolution is required in order to detect the small Doppler shifts associated with typical thermospheric winds. For example, a wind speed of 200ms$^{-1}$ detected from the 630nm red-line emission corresponds to a wavelength shift of $4.2\times10^{-13}$m (0.00042nm). For a 18.5mm etalon gap each bin corresponds to a shift of $4.19\times10^{-14}$m which corresponds to a wind speed of ∼43ms$^{-1}$. However, the FPI is capable of measuring fractional bin shifts so that accuracies of better than 5ms$^{-1}$ can be achieved.

The Atmospheric Physics Lab operates a network of FPIs in Scandinavia measuring the red and green atomic oxygen emissions (see section 1.7.1). There are three in the auroral oval: one each for green and red emission lines at Kiruna, Sweden and one for the red line at Sodankylä, Finland. There are two co-located FPIs in the polar cap at Svalbard. Each is capable of measuring both emissions: one is a standard FPI and one a Scanning Doppler Imager (see chapter 3).

## 2.2 Instrument Design

The general APL FPI design is shown in figure 2.1. It is mounted on an optical bench consisting of four parallel rods held in place with fixed brackets. Above this, built into the roof of the room, is a Plexiglass dome which allows sky-light through and is heated to prevent condensation and build up of snow.

Light is directed into the etalon by way of a tilted mirror which rotates to view different regions of sky. For the mainland FPIs, to take zenith measurements the mirror is rotated to point at a second mirror, which is pointed at the zenith, arranged in a periscope configuration. Whereas the Svalbard FPI has a mirror which, in one of the azimuth positions, is tilted to the vertical with a cam mechanism so that light from the zenith passes either side of it. The second mirror is, therefore, not required. For the mainland FPIs the off-zenith look directions are set to $45^o$ elevation whereas Svalbard is set to $30^o$. To inject calibration light, the source lamp illuminates a screen that scatters the light evenly across the field of view.

For the 240km altitude red line emission the $1^o$ field of view maps to a circular area on the sky of approximately 10km diameter. The FPI cycles through several viewing directions: north, east, south, west and zenith as well as a neon calibration lamp exposure. The Svalbard FPI also takes exposures at the south-east and south-west look directions pointed towards the auroral oval. Kiruna and Sodankylä have two common volumes: A (overlap to the north) and B (overlap to the south).

The mirror directs the light into the etalon, the most important component of the instrument (section 2.3). This is contained in an evacuated housing which is equipped with an electrically operated valve system and vacuum pump to vary the pressure. This enables precise tuning to optimise the peak position of the fringe pattern and also allows the pressure scanning for determining the instrument function (see section 2.4). In short,

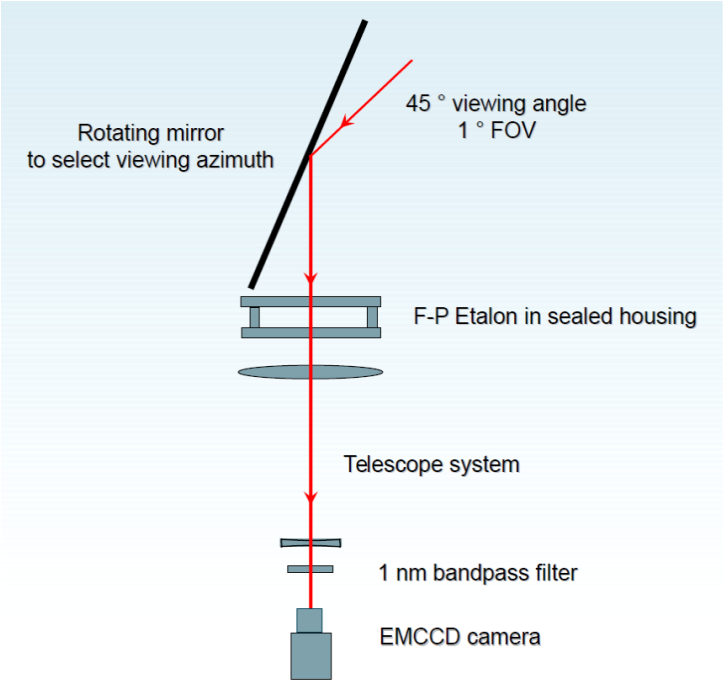


Figure 2.1: APL FPI set-up courtesy of Ian McWhirter.

the etalon creates the interference fringes as well as providing some preliminary filtering. The interference pattern is focused onto the detector using a Galilean telescope, i.e., a convex lens followed by a concave one. The effective focal length is 1.2m which includes the camera lens. Before the light reaches the detector it is passed through the filter. This is 1nm wide centred at 630.2nm. This removes the continuum from the incoming light to leave a bandwidth centred around the central wavelength. A bandwidth of 1nm is chosen to allow through all Doppler shifts of the emission line whilst minimising any other spectral contaminants of a similar wavelength. This is usually centred a little higher than the central wavelength, at 630.2-630.4nm, due to the filter response moving to a lower wavelength for off-axis light. The final destination for the incoming light is the EMCCD detector in the camera where the light is detected and an exposure is taken.

## 2.3 The Etalon

The etalon is the heart of the FPI. It is comprised of two partially reflective (80-90% reflectivity), perfectly flat parallel mirrors. Incoming light undergoes multiple reflections within the etalon cavity before emerging.

Suppose light of wavelength, $\lambda$, enters an ideal etalon with separation, d and refractive

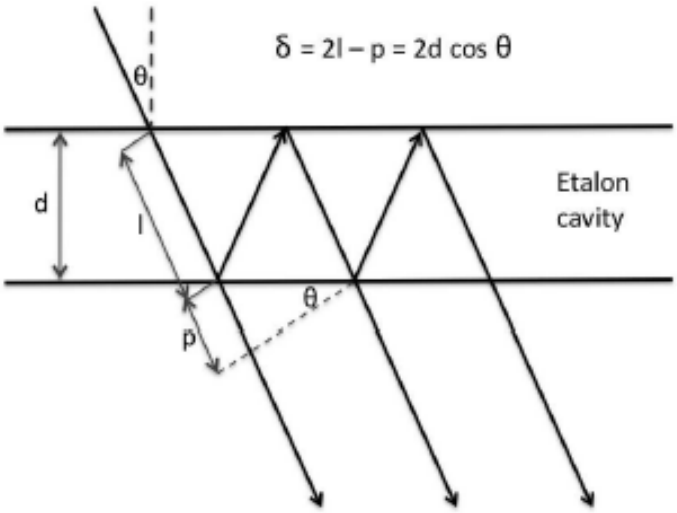


Figure 2.2: Light entering an etalon with a gap of d at incident angle $\theta$ and reflecting (and transmitting) multiple times within. Courtesy of McWhirter (1993).

index, $\mu$, at near normal angle of incidence, $\theta$, as shown in figure 2.2. The optical path difference between successive reflections will be:

$$\delta = 2dcos(\theta) \tag{2.1}$$

with a phase difference of:

$$\Delta\phi = \frac{2\pi\delta}{\lambda} = \frac{4\pi\mu d\cos(\theta)}{\lambda_0} \tag{2.2}$$

where $\lambda$ is the wavelength of the light in the etalon and $\lambda = \frac{\lambda_0}{\mu}$ where $\lambda_0$ represents the vacuum wavelength and $\mu$ is the refractive index. Constructive interference occurs when the optical path difference is equal to $n\lambda$ and destructive interference occurs when the optical path difference is equal to $(n + \frac{1}{2})\lambda$ for some integer n. The interference fringes produced are circular due to the optical path difference being dependent on $\theta$ so constructive interference will always be at a constant $\theta$. Figure 2.3 displays a typical interference pattern.

For a perfect etalon of reflectance R and transmittance T, the transmission, A(n), is given by the Airy function:

$$A(n) = \frac{1}{1 + \frac{4R}{(1-R)^2}sin^2(n\pi)} \tag{2.3}$$

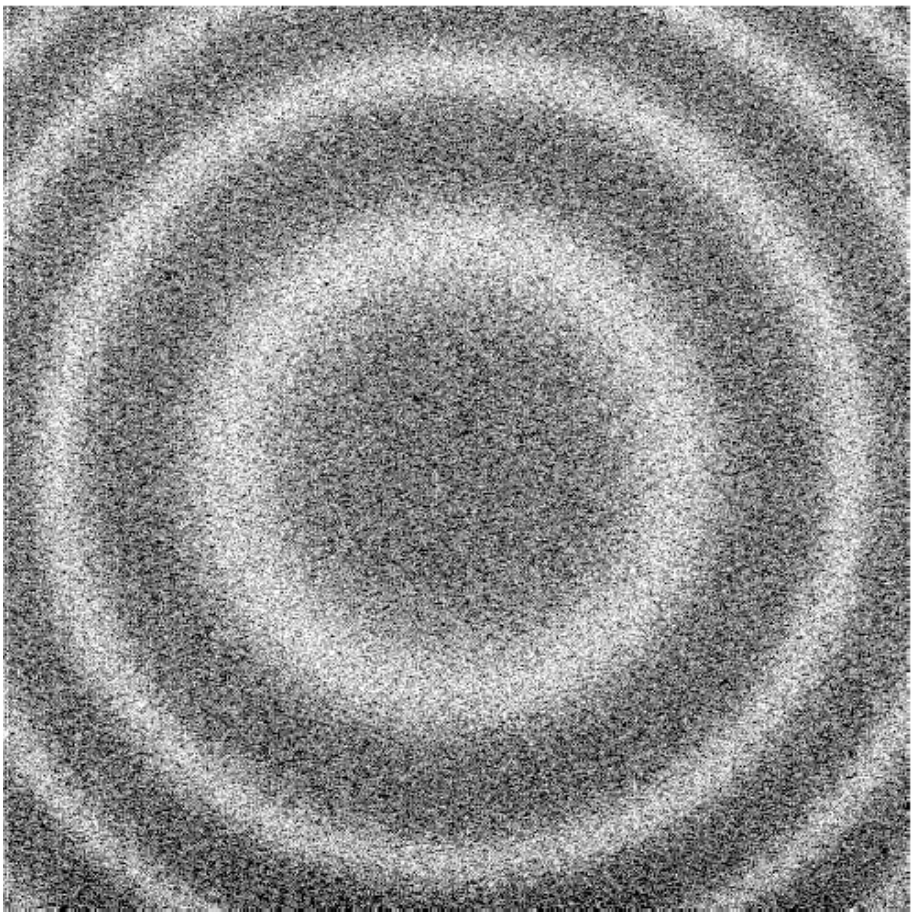

Figure 2.3: Typical FPI interference pattern.

Introducing the definition of coefficient of finesse which describes the reflectivity of the etalon plates:

$$F = \frac{4R}{(1-R)^2} \tag{2.4}$$

we can simplify the Airy function:

$$A(n) = \frac{1}{1 + Fsin^2\frac{\Delta\phi}{2}} \tag{2.5}$$

where $n\pi = \frac{\Delta\phi}{2}$. Maximum transmission occurs at integer values of n, given by:

$$n = \frac{2\mu d\cos\theta}{\lambda_0} \tag{2.6}$$

The free spectral range is the change in wavelength between each successive fringe

provided $\theta$, d and $\mu$ are constant. It is a key FPI parameter as it represents the range of wavelengths which can be displayed without overlap on to different orders. Equating the expressions for maximum transmission for order n and n+1 we get:

$$\frac{2dcos(\theta)}{\lambda} = \frac{2dcos(\theta)}{\lambda + \Delta\lambda} + 1 \tag{2.7}$$

which simplifies to:

$$\frac{\Delta\lambda}{\lambda(\lambda + \Delta\lambda)} = \frac{1}{2dcos(\theta)} \tag{2.8}$$

As $\Delta\lambda \ll \lambda$ and for near axis rays $\cos(\theta) \approx 1$, the equation yields the definition of the free spectral range:

$$\Delta\lambda_{FSR} = \frac{\lambda^2}{2d} \tag{2.9}$$

This shows that the free spectral range for the wavelength observed is defined by the plate separation, d. There is a trade-off when choosing gap size, as a smaller free spectral range will give greater accuracy however it must also be large enough to fit the whole spectral profile on the detector. Aliasing is avoided by choosing the free spectral range so that it is greater than the largest possible Doppler shift. The Svalbard FPI has a gap width of 18.5 mm and the Kiruna red line FPI has a gap width of 14mm which respectively corresponds to free spectral ranges of 11pm ($11\times10^{-12}$m) and 14pm. The finesse of the interferometer, not to be confused with the coefficient of finesse, is a useful quantity as it quantifies the sharpness of the fringes and hence the resolving power of the etalon. It is the ratio of the separation of adjacent fringes ($2\pi$) to the full width at half maximum (FWHM) of the fringe. At FWHM the intensity is half of, A(n), (equation 2.5) at maximum. By inspection it is clear that the FWHM of the airy function occurs at $Fsin^2\frac{\Delta\phi}{2} = 1$. The $\Delta\phi$ angle is sufficiently small that the sine term may be approximated by the angle and as half maximum occurs at $\pm\Delta\phi$, the FWHM is $2\Delta\phi$. Therefore the reflective finesse is defined as:

$$N_R = \frac{2\pi}{2\Delta\phi} = \pi\frac{R^{1/2}}{(1-R)} \tag{2.10}$$

## 2.4 Imaging and Scanning

In equations 2.3 and 2.6 we see there is an interdependence between $\lambda$, $\mu$, d and $\theta$. These relationships are necessary for two fundamental FPI operations: imaging and scanning. Imaging measures shifts in wavelength of a stationary interference pattern, by recording an image with fixed etalon gap and refractive index. Scanning measures the wavelength shifts by radially shifting the ring pattern through a fixed point detector. This is done by varying the etalon gap or refractive index. In the case of SCANDI, each pixel of the image can be considered a separate single point detector.

Both standard FPIs and SDIs utilise an imaging technique in order to measure a shift in wavelength of the incident light. It relies on the $cos\theta$ dependence to measure the change in angular diameter of a fringe, $\theta$. If a converging lens of focal length, f, is used to form the image of the rings, so the angular diameter can be related to the radial distance of the image, r, by the simple lens theory:

$$tan\theta = \frac{r}{f} \tag{2.11}$$

and since $\theta$ is small $\theta \approx \tan(\theta)$ so $\theta = \frac{r}{f}$. Therefore, the expansion for $\cos\theta$, ignoring higher orders is:

$$cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - ... \tag{2.12}$$

Therefore equation 2.6 can be written as:

$$n\lambda = 2d(1 - \frac{r^2}{2f^2}) \tag{2.13}$$

This shows that for small angles, there is a linear relationship between $n\lambda$ and $r^2$, i.e., adjacent maxima of fringes form a series of concentric circles and successive fringes have a constant separation in $r^2$ space.

The FPIs can measure an entire transmission profile for each of its $1^o$ FOV viewing directions, which consists of 2-3 orders of the radial interference pattern. With $1^o$ FOVs, there is little positional information in the rings therefore little is lost from integrating around each ring which also improves the signal-to-noise ratio.

The intensity values of the image pixels are added together according to the square

of their radial distance from this centre to give a data histogram in radius squared space, which as equation 2.13 has shown, is linear in wavelength space.

Another main functionality of an etalon is the ability to scan. The scanning method of measuring wavelength shifts uses the dependency of $cos\theta$ on:

- refractive index ($\mu$), used for creating laser profiles to derive neutral temperature data
- plate separation, d, used for SDIs to scan the sky (discussed further in section 3.3)

This method still involves imaging at the end.

## 2.5 Precision of Measurement

There are errors associated with spectroscopic measurements due to the finite width and shape of the emission line being studied (Hernandez, 1986). The measured signal, i.e., the flux of photons, follows a Poisson distribution (Parratt, 1961). Therefore the FPIs measure Poisson distributions, however, ones that are scaled by the quantum efficiency of the detector. If an emission line is sampled by a spectral scan at equidistant steps, then the centre of the emission line, $\bar{x}$, is:

$$\bar{x} = (\sum_{i=-\infty}^{\infty} x_i S_i)(\sum_{i=-\infty}^{\infty} S_i)^{-1} \tag{2.14}$$

where $x_i$ is the wavelength measurement at the ith step and $S_i$ is the number of events measured at that step. The variance of the determination of the centre of the line, $\sigma_x^2$, is given by:

$$\sigma_x^2 = (\sum_{i=-\infty}^{\infty} (x_i)^2 S_i)(\sum_{i=-\infty}^{\infty} S_i)^{-2} \tag{2.15}$$

Gagne et al (1974) show that for a Doppler broadened line of FWHM dg, the variance of determination of the centre of the line is given by:

$$\sigma_x^2 = dg^2 (N \ln(2))^{-1} \tag{2.16}$$

where N is the area under the spectral profile, i.e., total photon count. This relationship shows that to increase the accuracy of the determination of the centre of the spectral

profile (a step towards deriving the wind speed), reducing the line width (dg) is much more effective than increasing the number of photon counts (N).

Equation 2.16 shows that instrument broadening will decrease the precision of measurement. The bandwidth could be made narrower to reduce instrument broadening. However, if made too narrow, this could significantly reduce the number of photons allowed through the instrument and thus affect the resolving power of the instrument. This compromise between the instrument broadening and the incident flux is termed the Luminosity Resolution Product (LRP) criterion (Chabbal, 1953).

If both wind and temperature measurements are to be derived from the emission line, further consideration is required of the optimum FPI operation parameters. Hernandez (1986) shows that to produce the minimum uncertainty in one measurement would produce an increase in the uncertainty of the other. The LRP criterion is used to establish the choice of etalon aperture, etalon gap and plate reflectivity in order to meet the required compromise between resolution and luminosity, optimising for individual measurement parameters. However, if both wind and temperature parameters are taken into account, the optimum FPI operation parameters need to be chosen particularly carefully (Eoghan, 2000). This is a concern of our instrument designer and engineer, Dr Ian McWhirter.

## 2.6 Data Processing

After collecting the spectral data, it is processed in order to estimate the sky parameters: neutral wind velocity (from the line-of-sight Doppler shift), neutral temperatures and emission intensity. The line-of-sight Doppler shift is found from the angular diameter of the interference fringe. The neutral temperature and the emission intensity can be calculated from the full width at half maximum and the integrated area of the spectral line respectively. As the measured spectral data has a certain degree of noise, a function needs to be fitted in order to model the true sky backscatter function and to ascertain the best estimates of the spectral parameters (which define the function).

It should be noted that the Doppler effect is responsible for both temperature and wind velocity signatures in the spectra. An emitting species with a component of velocity along the line of sight away from (towards) the observer manifests as an apparent shift in wavelength, i.e., a red (blue) shift. Doppler shifts directly give the line of sight velocity from the change in peak position and indirectly give the temperature. Due to the random

thermal motions of the emitting species in all directions, the many small random Doppler shifts manifest themselves as an overall Doppler broadening of the emission.

The FPI data are stored in binary format as photon count profiles in wavelength space over 256 elements (or wavelength channels/bins). The first two fringes of the interference pattern are stored, i.e., 128 elements per fringe. For analysis, the first fringe of every profile is used because it is the sharpest and hence will introduce less uncertainty. It is sharper (than the fringes from the outer rings of the interference pattern) due to the outer rings being distributed increasingly closer together in r space (radial space) on the chip so that upon mapping to $r^2$ space (to get a constant radial distribution of rings), the outer rings are stretched radially, and thus, the resultant fringes are not so clearly defined.

### 2.6.1 Least Squares Best Fit

To estimate the sky parameters, a least-squares best fit is performed on each spectral profile (first fringe), in order to fit five parameters:

1. background continuum backscatter intensity
2. aerosol backscatter intensity (Mie scattering)
3. molecular species backscatter intensity (Rayleigh scattering)
4. neutral temperature, T (in wavelength units)
5. the Doppler shifted peak wavelength of the return signal, $\lambda_0$

Firstly, the fitting procedure involves modelling a discrete spectrum in wavelength space as $s(\lambda)$:

$$s(\lambda) = \sum_{j=0}^{2} a_j S_j(\lambda) \tag{2.17}$$

where $a_j$ are the backscatter intensities of the three functions, $S_j(\lambda)$, which contribute to the measured line, i.e., background continuum, aerosol (Mie) and molecular (Rayleigh) scattering, where T and $\lambda_0$ are parameters within the (Gaussian form) Rayleigh scattering function.

Secondly, the conventional $\chi^2$ parameter is used to describe the degree to which the model spectrum departs from the recorded spectrum, defined as:

$$\chi^2 = \sum_{n=0}^{N-1} \frac{(y_n - s_n)^2}{\sigma_n^2} \tag{2.18}$$

where $\sigma_n^2$ is the variance of the nth element of the observed spectrum, $y_n$, and N is the number of wavelength channels in the measured spectral line, i.e., 128. The best choice of model parameters yields the minimum $\chi^2$, so that:

$$\frac{\partial \chi^2}{\partial p_j} = 0 \tag{2.19}$$

for all j where $\{p_j\}$ is the set of five model parameters that require fitting.

Thirdly, a procedure needs to be implemented which searches for the optimum $\{p_j\}$ that minimises $\chi^2$. The procedure adopted involves using a combination of parameter optimisation strategies, both analytic and numerical methods. The numerical method used is called a 'grid search' which proceeds by minimising $\chi^2$ for one of the model parameters whilst holding all other parameters fixed. This is cyclically repeated for each model parameter in turn until there is satisfactory convergence of $\chi^2$. The analytic methods involve directly solving equation 2.19. The first is known as 'directly solving the normal equation' and the second is known as singular value decomposition (SVD). Another method, known as The Levenberg-Marquardt search (Press et al, 1986), is a numerical algorithm which uses as much analytic information as possible. It computes analytic expressions for the partial derivative of $\chi^2$ with respect to T and $\lambda_0$ and finds the steepest descent of these functions' gradients (i.e., a numerical search). The specific algorithms of these methods are beyond the scope of this work. For more detail on these procedures see Conde (1998).

There are various advantages and drawbacks of each parameter optimisation method mentioned above. The grid-search method is the most stable, however, it is very slow. The analytic methods are much faster, however, they are not capable of optimising for non-linear parameters within the model, i.e., T and $\lambda_0$. Furthermore, the normal equation method can be unstable if the spectra are noisy (as it cannot distinguish between the three backscatter functions), whereas the SVD approach is more stable. The Levenberg-Marquardt approach, similar to the grid search method, is capable of finding good estimates of the optimum non-linear parameters. Therefore, to most efficiently derive all the

spectral parameters, a combination of these methods is used. For more information on the specific combination, see Conde (1998).

### 2.6.2 Uncertainties in Fitted Quantities

The standard non-linear least-squares fitting technique (e.g., Bevington, 2003; Press et al, 1986) provides a generalised method for computing the variances of all the model parameters simultaneously, as follows.

The complete set of five variable parameters that require fitting (listed at the start of section 2.6.1) are denoted as $\{v_i\}$ where i varies from 0→5. We construct a complete 5×5 curvature matrix, $c$, encompassing all parameters using:

$$c_{jk} = \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} \frac{\partial s_n}{\partial v_j} \frac{\partial s_n}{\partial v_k} \tag{2.20}$$

where j and k are the row and column indices of c, N is the total number of spectral channels in the profile (i.e., 128), and $s_n$ corresponds to the nth spectral channel value of an analytic function of the form of equation 2.17. The parameter variances are then obtained directly from the diagonal elements of $c$ so that:

$$\sigma^2(v_1) = c_{11} \tag{2.21}$$

and we obtain variances for $\sigma^2(v_1) \rightarrow \sigma^2(v_5)$ for each of the five fitted parameters.

For the FPI's measured spectra, the resultant uncertainties in the fitted useful sky parameters typically are, for:

- the neutral temperature, ∼90K
- the wind speed, ∼10ms$^{-1}$
- the emission intensity, i.e., from Rayleigh scattering, ∼30 counts (estimate is for recent years' intensity software settings and state of optics, see section 2.7)

### 2.6.3 Signal-to-Noise ratio

The signal-to-noise ratio (SNR) is another useful quantity to analyse the quality of the data. It is calculated from the power spectrum of the sky profile (i.e., the squared spectral profile), and is defined as the power spectrum's fundamental component's power (height

of the main peak) divided by the average value of it's high frequency, noise-dominated components' power (average height of remaining peaks on power spectrum after the fundamental component, i.e., main peak, is removed). Typical SNRs are ∼3500 and can reach as high as the order of $10^6$. Generally, for these spectra, a 'good' value is thought to be roughly >2500. This threshold value would be regarded as a very strong SNR in other scientific domains. However, when regarding FPI spectra, this is not the case. The signal needs to be particularly strong with FPI spectra as a slight shift in peak bin number translates to a large change in Doppler shift (∼43ms$^{-1}$ per bin).

### 2.6.4 Line of Sight Velocity

To derive the line of sight winds, the shift in peak position (given in bin number units by the spectral fitting algorithm), needs to be translated into a shift in wavelength before it is converted into a velocity value. Starting with the Doppler shift equation for an emitting species:

$$\frac{v_x}{c} = -\frac{\Delta\lambda}{\lambda_0} = \frac{\Delta\nu}{\nu_0} \tag{2.22}$$

where $v_x$ is the LOS velocity, c is the speed of light, $\lambda_0$ is the wavelength from the emitting species at rest, $\Delta\lambda$ is the measured apparent shift in wavelength away from $\lambda_0$, and $\nu$ is the corresponding frequency. The negative sign next to the wavelength quotient explains how the measured wavelength becomes shorter for a positive LOS velocity (emitting source moving towards the observer, i.e., blue shift) and becomes longer for a negative LOS velocity (red shift).

By substituting the FSR equation derived previously in section 2.3, $\Delta\lambda_{FSR} = \frac{\lambda_0^2}{2d}$, we get the expression for the velocity change for a single bin difference:

$$v_{bin} = \frac{1}{n_{bin}}.\frac{\lambda_{sky}c}{2d} \tag{2.23}$$

where $\Delta\lambda$ from equation 2.22 is $\frac{\Delta\lambda_{FSR}}{n_{bin}}$ and $n_{bin}$ is the number of bins in a FSR. The value for $n_{bin}$ can vary slightly from exposure to exposure so is recalculated every exposure. Within the raw data analysis software the FSR for the particular spectral profile is calculated by measuring the bin difference between the first and second order fringes. Typically it is found to be equal to 117 bins for the Svalbard red line FPI with an 18.5mm

gap. This gives a velocity of about 44ms$^{-1}$ per bin. This particular value is calculated for each spectral profile within the raw data analysis software and so the relative LOS wind velocities can be calculated. However, these values need further processing. They need to be corrected for any instrumental temperature drift and a baseline needs to be found to estimate their absolute values.

The calibration lamp is used to monitor and compensate for any instrumentation drift. Although the etalon housing and optical bench are temperature stabilised, there can still be small fluctuations which minutely affect physical dimensions of any part of the instrument and the air pressure inside the etalon chamber. The calibration lamp data are used to correct the sky data for these variations. Occasionally there may be drifts due to a leak developing in the etalon chamber during the long periods for which the instrument is unattended. This again can be fully compensated for.

The zero velocity reference is calculated by taking the difference between the zenith and calibration lamp peaks throughout the night. The average is taken and is called the offset which is an estimation of the shift required to derive absolute wind values. This step uses the assumption that the vertical wind throughout a night averages to zero, with a possible error of up to $\pm$10ms$^{-1}$ (Aruliah et al, 1991). This is a reasonable approximation as the vertical winds are typically an order of magnitude smaller than the horizontal. Then the zero Doppler shift baseline at 15 minutes is calculated by adding the offset to the calibration lamp peaks. This step corrects for the instrument temperature fluctuations. The baseline is then subtracted from all wind measurements as an overall correction.

### 2.6.5 Temperature

The derivation of the FPI neutral temperature measurements from the FWHM of the spectral profiles is shown below. Kinetic theory relates average kinetic energy and temperature of a fluid as:

$$\frac{1}{2}m\overline{v^2} = \frac{3}{2}kT \tag{2.24}$$

where m is the mass of an atom, $\overline{v^2}$ is the mean squared velocity of the atoms where we are in the bulk rest frame of the gas, k is the Boltzmann constant and T is the absolute temperature. The velocities of the atoms in a fluid may also be described by the Maxwell-

Boltzmann distribution:

$$\frac{dN(v_x)}{N} = \frac{\alpha}{\sqrt{\pi}} \exp\left(-\frac{v_x^2}{\alpha^2}\right) dv_x \tag{2.25}$$

where $v_x$ is the LOS velocity, $\frac{dN(v_x)}{N}$ is the fraction of atoms with velocity between $v_x$ and $v_x + dv_x$, and $\alpha$ is the most probably velocity given by:

$$\alpha = \left(\frac{2RT}{M}\right)^{\frac{1}{2}} = \left(\frac{2kT}{m}\right)^{\frac{1}{2}} \tag{2.26}$$

where M is the relative mass number for the species and R is the universal gas constant. By substituting equation 2.22 into equation 2.25, $v_x$ can be expressed in terms of frequency. The fraction of atoms emitting in the frequency range $\nu$ to $\nu + d\nu$ is given by:

$$\frac{dN(\nu)}{N} = \frac{\alpha}{\sqrt{\pi}} \exp\left(-\frac{c^2(\Delta\nu)^2}{\nu_0^2\alpha^2}\right) \frac{c}{\nu_0} d\nu \tag{2.27}$$

The intensity of emission at frequency $\nu$, $I_\nu$, is proportional to the number of atoms with frequency $\nu$, $dN(\nu)$, so the intensity per unit frequency interval (for the line profile) can be expressed as follows:

$$I_\nu = I_0 \exp\left(-\frac{c^2(\nu_0 - \nu)^2}{\nu_0^2\alpha^2}\right) \tag{2.28}$$

where $I_0$ is the peak intensity. The values of frequency at which the intensity is half of that at the centre, denoted $\nu_{\frac{1}{2}}$, will have $I_\nu = \frac{I_0}{2}$. Substituting this into equation 2.28 and taking the logarithm of both sides gives:

$$\frac{-c^2(\nu_0 - \nu_{\frac{1}{2}})^2}{\nu_0^2\alpha^2} = \ln\left(\frac{1}{2}\right) \tag{2.29}$$

which leads to the FWHM expression:

$$\delta\nu_D = \frac{2\nu_0}{c}\sqrt{\frac{2RT\ln 2}{M}} = 7.16 \times 10^{-7}\sqrt{\left(\frac{T}{M}\right)}\nu_0 \tag{2.30}$$

and in dimensionless form:

$$\frac{\delta\nu_D}{\nu_0} = \frac{\delta\lambda_D}{\lambda_0} = 7.16 \times 10^{-7}\sqrt{\frac{T}{M}} \tag{2.31}$$

Thus, the temperature is related to the FWHM, and by using this relation, the thermospheric temperature can be deduced from the emission profile. However with real etalons there will be imperfections within the optical path, such as departure from flatness of the plates, which will act to broaden the measured spectra (section 3.9.2 goes into more detail about this phenomenon).

The additional instrumental broadening is called the instrument function and this needs to be removed before the temperatures are calculated. The instrument function is measured with HeNe laser light (632.8nm) and is subsequently deconvolved with the measured spectra to leave behind pure sky emission spectra. Laser light is necessary as it is so spectrally narrow (i.e., so spectrally pure that the spectral profile is very narrow), it is the closest light source to a delta function (although it is actually a slim Airy function). Therefore, when imaging the laser line with the FPI any width it has accumulated will be mostly due to broadening inherent to the instrument.

A series of laser profiles are created by performing a pressure scan. This is a process whereby the etalon is completely evacuated and then small amounts of dry, filtered air are introduced to increase the value of $\mu$ in steps. A laser interferogram is taken at each step, so a series of instrument functions with the peak of the profile located at every radial distance is recorded. The profile with the peak position closest to that of the subsequently recorded sky peaks is then taken as the instrument function. Pressure scans are executed at the beginning and end of every season to record any minute changes in the FPI optics which alter the instrument function. Typical drifts in the instrument function can be as large as $\sim$2 bins per day.

### 2.6.6 Intensity

The emission intensity is calculated from the integrated area under the spectral line so can be inferred directly from the spectral fit. The intensity data are relative measurements. The reason for this is explained further in section 2.7. The data can be cross-calibrated against an instrument with an absolute measure of atomic oxygen red or green line emission intensity such as, for the Svalbard FPI, the Meridian Scanning Photometer. A demonstration of this can be found in section 5.5.6.

## 2.7 Intensity Variability

The illumination recorded by the camera after the light has passed through the entire instrument gives an indication of relative emission intensity. However, over the years, the instrument design and the camera type have both evolved considerably so comparisons on this time scale are difficult. Furthermore, the recorded intensity is affected by several parameters. Examples are changes in EMCCD gain due to ageing of the gain register and drifting of the response of the interference filter. Furthermore, the instrument is unattended for long periods and so the optics and dome cannot be cleaned as often as would be desirable. The calibration lamp is also not a reliable source of uniform illumination as these dim with age and then have to be replaced.

The calibration lamp cannot be used to monitor the detected illumination; they are operated until they grow dim and fail and then they are replaced with new brighter ones. Therefore, a better measure of detected illumination is the 630nm sky light. However, when analysing this source of emission, we must be cautious of long-term sky emission intensity trends, which we may expect to be periodic with the solar cycle.

The average intensity of the 630nm detected intensity is plotted for 1981-2013 for both KEOPS and Svalbard in figure 2.4. Kiruna's data are in the top three plots and Svalbard's are in the bottom three. For each site, the first plot shows the average yearly detected intensities. There seems to be no atmospheric trend in the data so we may assume the trend is purely due to changes in the instrument set-up. The third plot is the number of data points included in each yearly average which ranges from 78-23401 for Kiruna and 84-87747 for Svalbard. It shows the statistical samples to be much larger in the more recent years due to the ability to reduce integration times. The error bars show the standard deviation for each yearly bin.

For Kiruna, the detected intensities were relatively low from 1982-1999, whereas Svalbard's remained low until 2005. Both sites' detected illumination dramatically increased with Kiruna's reaching a maximum in 2003 and Svalbard reaching a maximum in 2007. Both sites reverted to relatively small magnitudes by 2010. For the Svalbard FPI, the EMCCD camera was installed into the FPI in 2005, resulting in changes in software which increased intensity values. In 2007 the Svalbard FPI was moved to a new location (at the KHO) and its hardware and software were reset explaining the dramatic intensity increase. In 2009, a year of green-line measurements were taken so the software was altered to re-

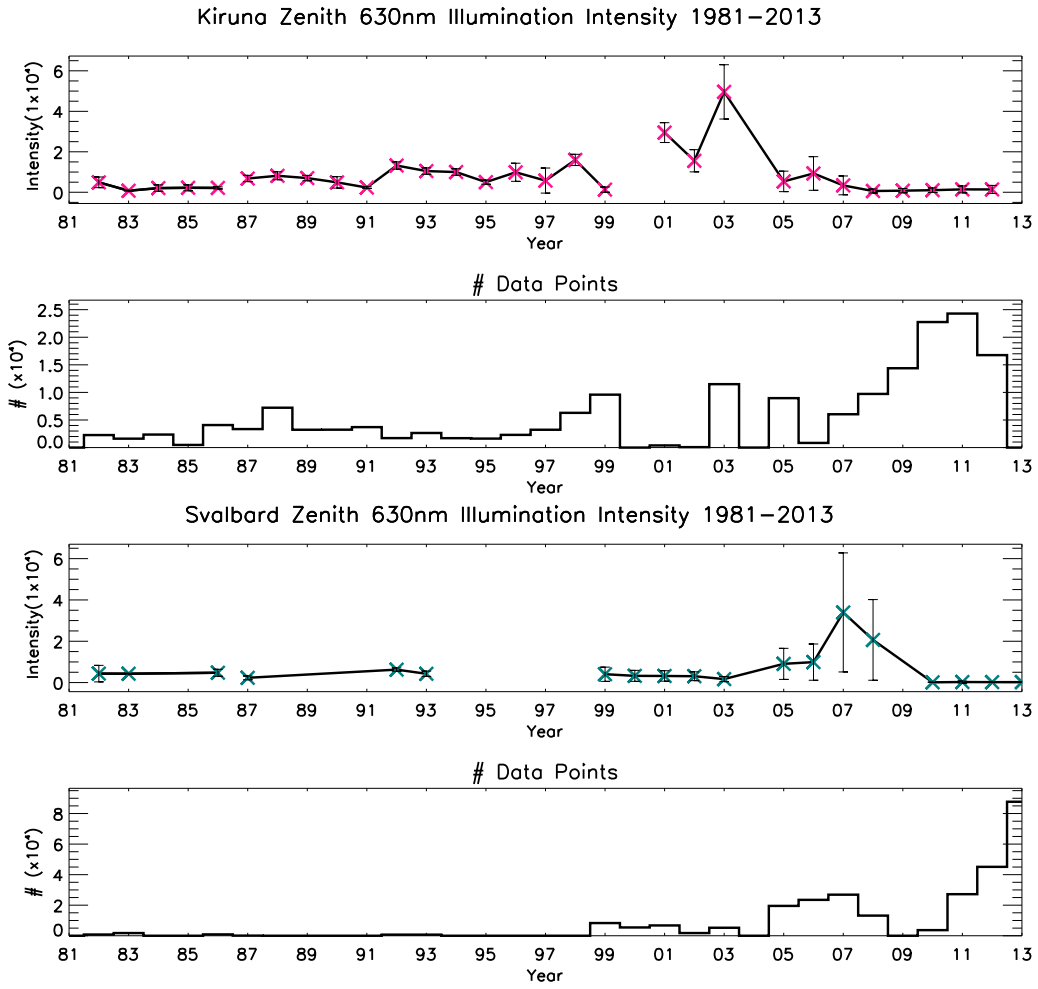


Figure 2.4: Sky light (630nm) illumination intensity variation from the zenith look direction over 1981 to 2013 for Kiruna (top 3 plots) and Svalbard (bottom 3 plots). For each site the top plot shows all the data, the middle plot shows the low intensity data and the bottom plot shows the number of data points included in each average. The error bars show the standard deviation of each yearly bin's average value plotted.

scale the brighter intensities to more moderate values. These software settings were not reset for the following year when the red-line emissions were measured again, resulting in the smaller intensity values. These software resettings are likely to explain the relatively large standard deviation in the 2007-2008 bins. These bins will share intensity values from before and after the date that the software was reset (mid-season), which results in a mixture of high and low relative intensities.

Overall this plot demonstrates the large long-term variability of the instrument sensitivity. It is useful as a reference to see which years' intensity values can be meaningfully cross-compared.

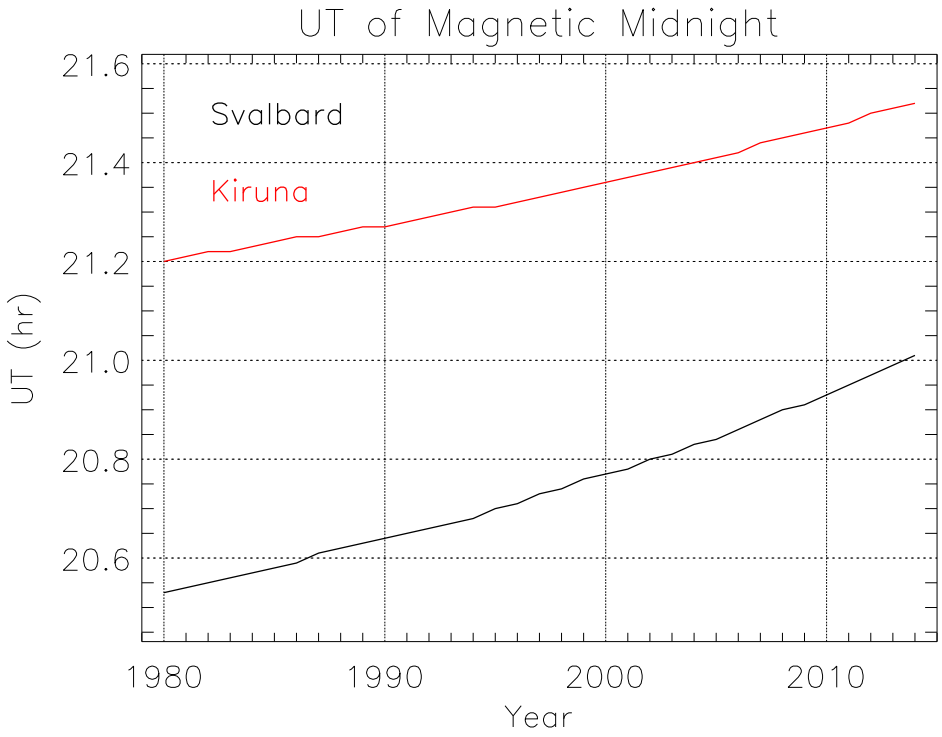


Figure 2.5: Svalbard and Kiruna's magnetic midnight variability from 1980 to 2014 as the magnetic pole drifts using IGRF data.

### 2.7.1 MLT change 1980 to 2013

As the magnetic pole drifts around the polar cap, the location of the magnetic pole with respect to our observing sites changes. This results in the time of magnetic midnight getting increasingly later throughout the years. This is demonstrated in figure 2.5 which uses the International Geomagnetic Reference Field (IGRF) model to calculate the universal time of magnetic midnight from 1980 to 2010 for both our observing sites; Svalbard is in black and Kiruna in red. Kiruna's magnetic midnight has drifted 20 minutes in the 35 year stretch from 21:12UT to 21:32UT whereas Svalbard has drifted 26 minutes from 20:32UT to 21:06UT; not significant amounts. This is a useful reference to use when analysing the data in magnetic time coordinates.

This page was intentionally left blank

# Chapter 3

# SCANDI

The Svalbard SCANing Doppler Imager (SCANDI) was deployed first at the Auroral station Longyearbyen, Svalbard ($78^o$N $16^o$E) in December 2006. It was moved to the new Kjell Henriksen Observatory (KHO) in November 2007.

SCANDI is an all-sky FPI, based on a novel design developed by Conde and Smith (1997), which uses a variable width etalon to enable a scan of the whole sky. It is capable of simultaneously measuring spectra over several tens of locations within a $150^o$ FOV, i.e., $\sim$1200km ($\sim$600km) radius circular area of the sky at $\sim$240km ($\sim$120km) altitude for the red 630nm emission (for the green 557.7nm emission).

The Atmospheric Physics Laboratory (APL), UCL, built and currently maintain SCANDI. The design of SCANDI's optical hardware and mechanical engineering were performed by Dr Ian McWhirter and Dr Andy Charalambous. The instrumentation control and collection software were written by Dr Ian McWhirter, who also provides continuous support on the day-to-day running of the instruments. The capacitance-stabilised etalon was developed by Dr Ian McWhirter (1993) and is a critical component of SCANDI for maintaining the accuracy of the variable etalon gap width.

The original SDI spectral fitting software, written by Prof Mark Conde, was adapted to the SCANDI instrument by Dr Eoghan Griffin. There are two SCANDI quick-look data plotting programs written by Dr Iris Yiu. Firstly, there is the horizontal wind-fit sky-map which uses an algorithm formulated by Conde and Smith (1998). Secondly, there is the sky-map plot of the derived sky parameters (temperature, LOS wind and emission

intensity). Dr Anasuya Aruliah, Dr Ian McWhirter and I are currently responsible for monitoring the operation of the instruments. I am responsible for the data transfer and the current SCANDI data analysis.

My other instrument-related responsibilities, which will be described in the coming chapter, include further adaptation of the horizontal wind vector analysis from Conde (2002) and an upgrading of the 25-zone map to a 51, 61 and 91-zone map. I also am responsible for the final calibrations of the data, including an annular intensity correction and correct baseline determination. I was directly involved in two experimental campaigns (CUSPN and CUSPII), as well as two remote experiments (the auroral upwelling experiment and solar eclipse experiment). I participated in one instrumentation maintenance field trip where optical re-alignment and various calibrations were carried out.

In the following chapter my developments on the existing SCANDI infrastructure will be discussed. At the end of the chapter horizontal wind climatologies will be presented with corresponding SuperDARN (Chisham et al, 2007) data to verify the dataset.

## 3.1 Introduction

SCANDI (Aruliah et al, 2010) is a spectrometer capable of simultaneously measuring the spectra over several tens of locations within a $150^o$ FOV, which corresponds to a circular region of the sky ∼1200km in diameter (at 240km altitude). Its design is based on the innovative concept of Conde and Smith's (1997) SDI which was installed at Poker Flat, Alaska. SCANDI, however, has an unprecedentedly high latitude location for an SDI. It is strategically placed at the polar cap site of Longyearbyen, so that it can observe the open-closed field-line boundary and more specifically the cusp region, an active area of research within the APL (Yiu, 2012). Its high spatial resolution, ∼100km, allows mesoscale studies of the thermosphere which are now understood to have a significant impact on the upper atmospheric dynamics (Aruliah et al, 2009, 2010; Griffin et al, 2008, 2009; and Kosch et al, 2010). It has the capability of measuring both the red and green line atomic oxygen emissions and so is useful for E and F region comparisons. Its FOV overlaps with the EISCAT and SuperDARN radars so can also be used for ion-neutral interaction studies (see section 3.11.1 and chapter 7). Its data has a time resolution of 7 minutes and from November to January SCANDI operates continuously 24 hours a day due to the dark winter polar cap.

SCANDI's FOV is split into annular rings about the zenith and then each ring is split azimuthally into equally spaced sectors to make a particular zone pattern (see figure 3.2). Each pixel on the EMCCD detector records spectra and the chip is sectioned into zones whereby the spectra within each zone are averaged. SCANDI's zone map's spatial resolution was increased from 25 to 61 zones in preparation for the onset of solar maximum and in particular for the CUSPN campaign in Svalbard in January 2012. SCANDI's several data analysis programs were altered for the new zone pattern. It was necessary to modify the spectral-fitting, line plot, map plot and horizontal wind fitting programs which will be shown further along in this chapter.

## 3.2 Instrument Design

SCANDI is a seperation-scanned, all-sky (wide-field) imaging spectrometer of the type described by Conde and Smith (1997). Figure 3.1 shows a 3D (left) and a 2D (right) schematic diagram of the instrument set-up. The instrument is 3 metres long and its optics are mounted on a cylindrical construction. The instrument is temperature controlled in order to maintain image stability. A fish-eye lens accepts light from a large ($\sim 150^o$) FOV of the sky.

The primary image is projected through the collimator which directs the incoming light into a 50mm diameter light path into the filter. The filter removes any background continuum of light and can be moved by the control software into four different positions to allow different bandwidths through the optics. Two positions hold the airglow filters (1nm width) centred at 630.2nm (red line) and 557.9nm (green line). Both are centred 0.2nm higher than emission peak to suppress reduction of transmission for off-axis rays. The third filter is centred at 632.8nm (10nm width) for the laser calibration light. The fourth is left empty to allow broad-band imaging. After filtering, the emerging light beam is then expanded to 150mm diameter to match the etalon aperture. The etalon is capacitance stabilised (see chapter 11, McWhirter (1993)), i.e., it is spaced by three piezoelectric electric transducers (PZTs) which change size with applied voltage. Furthermore, the cavity gap is monitored and fine-tuned with a capacitance sensor which makes small adjustments to the piezoelectric voltage. This functionality varies the angular diameter of the interference pattern which allows the sky to be scanned (see section 3.3).

Light emerging from the etalon is passed through a Galilean beam reducer to a com-

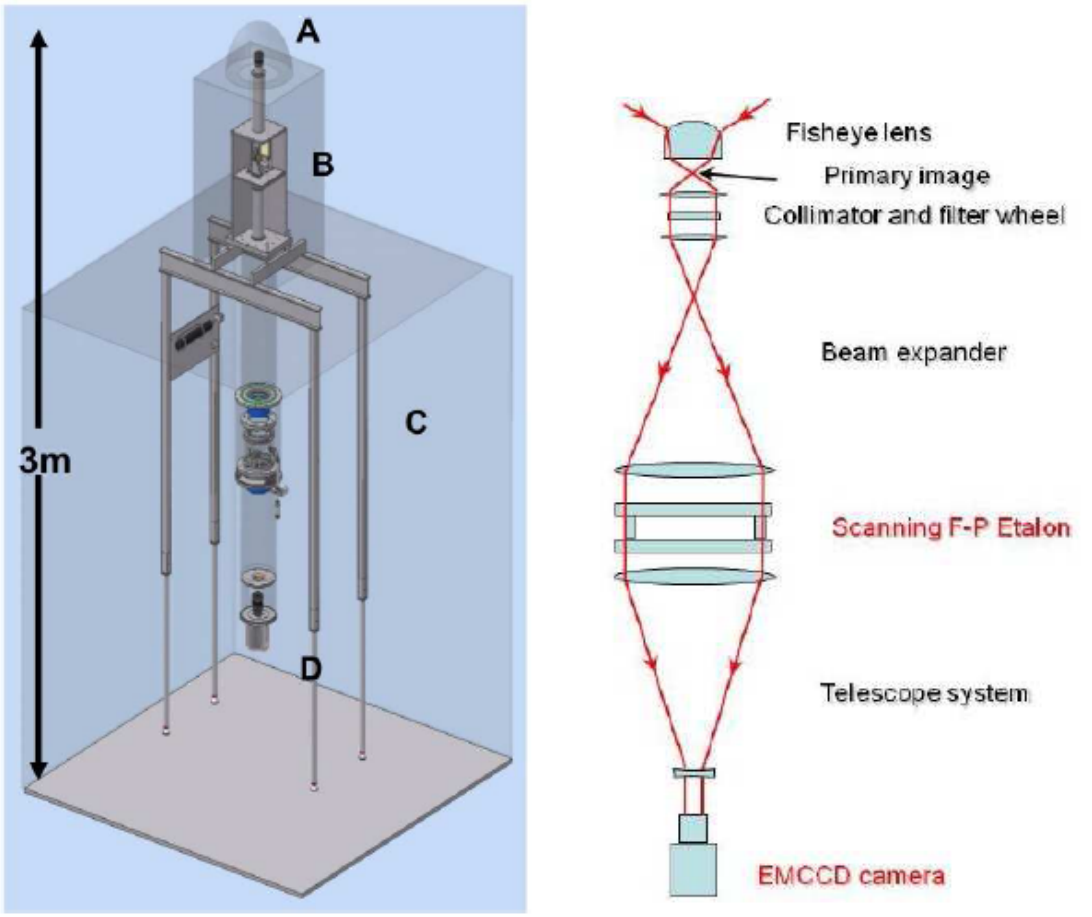


Figure 3.1: This figure shows the three-dimensional (left) and two-dimensional (right) schematics of SCANDI (courtesy Aruliah et al., McWhirter, I., and Charalambous, A., 2010). SCANDI is about 3 metres long, and for the left diagram, the labels are: A. All sky fish-eye lens, B. Laser and calibration lamp panel and filter wheel, C. Capacitance-stabilised scanning etalon, D. EMCCD, Andor iXon camera. The diagram on the right shows the simplified optics (not-to-scale).

mercial camera lens (f/1.8) which focuses the final image onto the Electron Multiplying CCD chip (EMCCD). The EMCCD has a quantum efficiency of 90%, substantially larger than the best Imaging Photon Detectors (IPDs) of the 1980s which typically achieved 15% (McWhirter et al, 1982). The EMCCD's read-out noise is negligible due to its electron multiplying functionality, i.e., the signal carrying electrons pass along the gain register (a special row in the chip lattice), additional electrons are produced by impact ionisation (of the signal-carrying electrons), thereby amplifying the signal. This process also amplifies the dark current, so to reduce this, there is a deep cooling of the EMCCD maintained with a Peltier thermoelectric cooler. Heat is removed from the hot side of the Peltier junc-

tion by water and fans which cool the chip to ∼-90$^o$C. The amplified signal and reduced noise ensure an extremely high signal-to-noise ratio (typical values of 4000) necessary for successful SDI operation (see section 3.7).

SCANDI's optics are comprised of two main systems: the FPI and the all-sky imager. Each of these must be independently aligned and focused. The laser is used to do this for the FPI. The light is injected through the appropriate laser filter to give stable fringes. The final camera lens is adjusted to focus the fringes and the fringe pattern is centred by adjusting the three etalon supports. The all-sky imager however is focused with broad-band light allowing the sky image to be centred and focused.

## 3.3 Scanning

Sky scanning is performed by varying the etalon spacing, d, over n discrete steps (through a full FSR) while $\mu$ remains constant (as mentioned in section 2.4 in chapter 2 on the narrow-field FPI). This is possible as photons detected at different pixels represent different axial angles of incoming light. Therefore, spectral profiles built up at each pixel can be mapped to different positions in the sky. The scanning process causes the interference pattern to shrink or expand radially. Stepping the interference pattern across an entire FSR means that each pixel, at one of the steps in the scan, experiences a fringe maximum. In fact, after a complete scan each pixel has detected photons for every spectral phase channel so can build an entire spectral profile. However since pixels at different radial locations on the chip will not share the same order of interference at the same scanning step (see equation $n\lambda = 2\mu d\cos(\theta)$), all pixels profiles are out of phase from each other. Therefore, all pixels of the detector need to undergo a process called phase mapping in order to decode their offset phases. This then yields final spectral profiles which are in phase with each other and eventually, from which, relative velocities may be inferred.

## 3.4 Phase Mapping

Phase mapping ensures all acquired spectral profiles are in phase with each other. The main tool used is the phase map. This is essentially a reference table which is looked-up at each step of a sky scan for each pixel, which provides the information that places the incident photon into the correct phase channel of a spectral profile. A phase map is created using a 632.8 nm He-Ne laser source for a fixed velocity reference line. By passing this

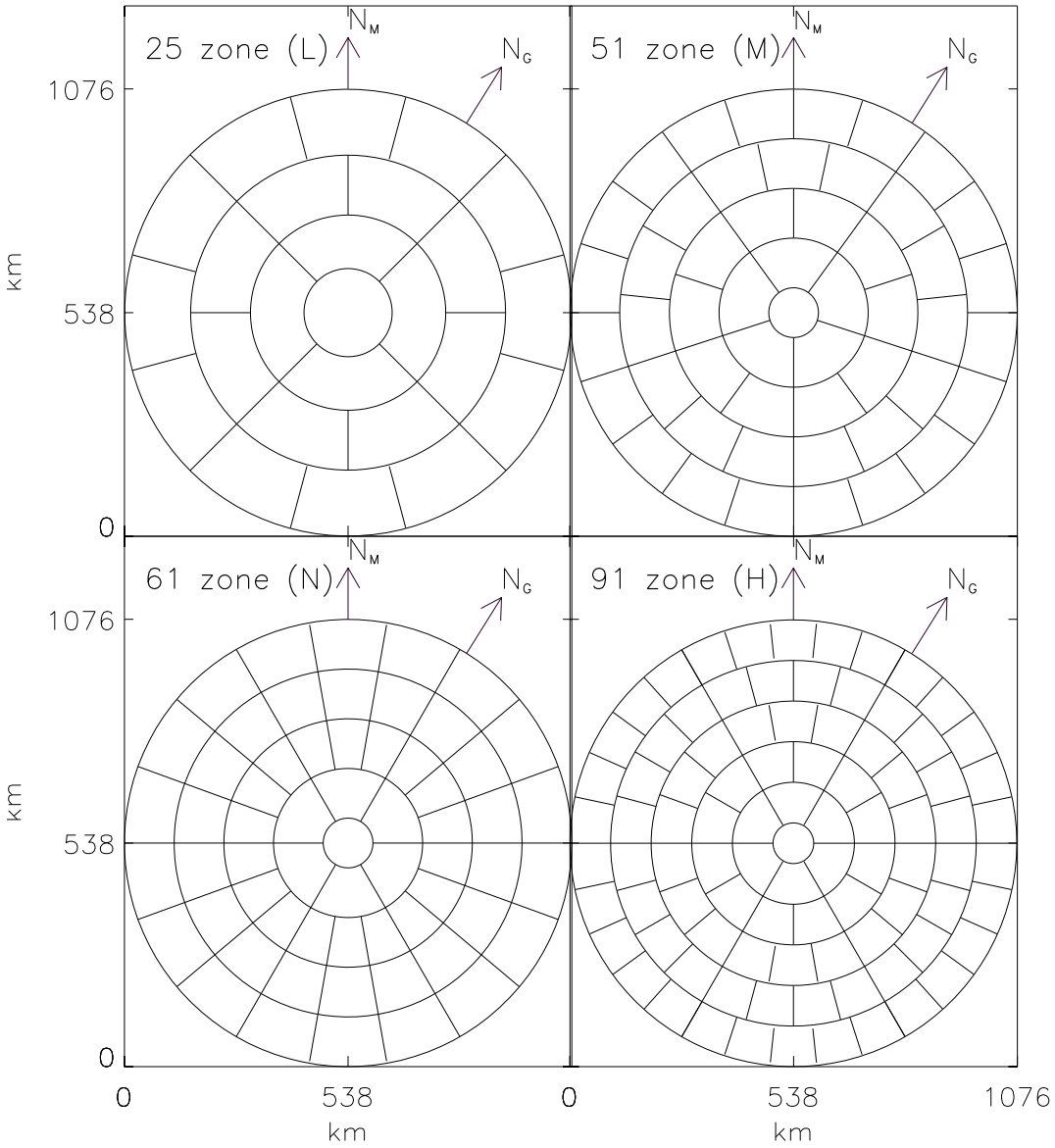


Figure 3.2: The four zone configurations as sky projections (in km) that are currently available in SCANDI's software: 25 zone as low resolution labelled 'L', 51 zone as medium resolution labelled 'M', 61 zone as new resolution labelled 'N' and 91 zone as high resolution labelled 'H'. These labels are used as the APL's SCANDI file naming convention.

light through the instrument and scanning, the scanning step at which a stationary source (the laser light) peaks at each pixel may be ascertained. Before recording to the phase map, in the instrument control software, the laser transmission function is scaled-down to cycle over the same FSR as the sky. The FSR of the laser light is scanned in 120 etalon steps. This is re-scaled to the 119 (115) steps required to scan the FSR of the red line (green line) so that the transmission functions exactly overlay.

Due to this phase mapping process the resultant spectral profiles for each pixel are shifted with respect to the scaled laser light, i.e., they are centred about 632.8nm bin position rather than the 630nm (or 557.7nm) bin position. This offset is removed which is discussed further in section 3.6.

## 3.5 Zone mapping

Generally, a single pixel will not detect enough photons to achieve sufficient signal-to-noise ratio to produce a useful spectral profile in a reasonable exposure time. Therefore spectra from neighbouring pixels are summed in order to form one spectral profile with an adequate signal-to-noise ratio. This is valid as, due to the phase mapping process, all the pixels' spectra are in phase so may be added. Theoretically all pixels' spectra could be added to create one grand profile representing the whole 1200km across field-of-view (FOV) however this would not be taking advantage of SCANDI's mesoscale measurement potential. Therefore, to retain directional information and some spatial resolution, the pixels are divided up into zones which map onto the sky FOV. There are four different zone maps currently available for SCANDI displayed in figure 3.2. The FOV can be split into a 25, 51, 61 or 91 zone configuration, and one of these is chosen with regard to geomagnetic activity levels whereby increased auroral emissions would allow a higher resolution map to be implemented. These zone maps will be explained in further detail in section 3.8.

The sky zone map is related to the chip zone map subject to the all-sky lens projection. An attempt can be made to map the incoming light rays which are projected through the fish-eye lens and then the instrument optics before reaching the detector. Firstly, the equi-solid-angle fish-eye lens projects the rays, from a $150^o$ FOV, into an image defined

by the ideal fish-eye lens model:

$$r = 2f \sin(\frac{\theta}{2}) \tag{3.1}$$

where r is the radial distance between a point on the image and the optical axis, f is the fish-eye focal length and $\theta$ is the angle between the incoming ray and the optical axis (AKA the zenith). However with a real lens the projection deviates a few percent away from this model due to imperfections in the lens. In addition, the remaining instrumental optics may further distort the projected image. Therefore it is not viable to project the image theoretically. Instead, the star field is used to calibrate the instrument zone positions with the sky. The method is described in (Yiu, 2012). She used the azimuthal distribution of stars to centre the instrument and orient it correctly azimuthally. She performed a least squared fit between the observed star field and the field generated by the SkyMap program to give a true projection model derived in the format of equation 3.1:

$$r = 2.28 \sin(0.46\theta) \tag{3.2}$$

This is the definitive equation used to map an incoming ray from the zenith angle, $\theta$, direction onto a radial distance on the chip. However this does not define the exact region of the sky from which the ray originates. To ascertain the emission source location, the assumption that the emission altitude lies at 250km is made. Even though this is not a fixed altitude in reality and the emission layer is known to expand and contract with the pressure gradient, it is a reasonable assumption as there is a large thermospheric scale height (of ∼50km) and the full width half maximum of the layer is also wide, estimated to be ∼50km (Vlasov et al, 2005). Therefore, using basic trigonometry, the zenith angle of incoming light is related to the radial distance on the sky by:

$$R_{250} = h \tan(\theta) \tag{3.3}$$

where $R_{250}$ is the radial distance from the point vertically above the observatory at 250km altitude, h is the emission height (250km) and $\theta$ is the zenith angle of the incoming light (zenith angle is measured from the vertical (zenith) direction). There is now enough information to map any point in the sky to the observatory.

## 3.6 Spectral Fitting

With a phase and zone map defined, spectral accumulation may commence. The acquired spectra are stored in binary format with intensity values recorded for each spectral bin. These data are analysed using a program which fits model Gaussians to the spectra and translates the peak positions into wind speeds and the peak widths into temperatures, finally outputting the data in ascii format. The same least-squares best fit procedure is applied as for the narrow-field FPIs. This is described in more detail in section 2.6 in chapter two. The fitting procedure and subsequent data processing (see section 2.6.5 and 2.6.4) yield estimates of relative neutral wind, absolute neutral temperature and relative emission intensity.

There is a possibility that a different fitting function would give a more accurate determination of the peak position, e.g., an Airy, Voigt or even polynomial function. Comparisons between the different fitting functions is a project that needs to be investigated within the APL group. Gaussian functions are used at the moment because they are a reasonable fit to the spectra, they are easy to implement and do not require much computing power (which was a concern when the APL's fitting algorithms were written in the 1980s). Another useful aspect of using a Gaussian fitting function is that the Maxwell-Boltzmann distribution can be easily manipulated into a Gaussian form with a temperature value corresponding to the FWHM (see equation 2.25 in chapter 2). Therefore, regarding temperature value estimations, the Gaussian is the best function to fit to the spectra, however, regarding the wind value derivations, there may (or may not) be a better fitting function to use, although the Gaussian function is adequate.

The phase mapping and subsequent data processing measure relative Doppler shifts due to the use of a non-sky-wavelength calibration source in the phase mapping (He-Ne laser at 632.8nm). This causes the derived velocities to be relative to the zero-velocity of the laser line manifesting as an essentially unmeasurable Doppler shift offset. In theory, this offset could be calculated by observing the bin position of the scaled laser profile peak and calculating the number of complete FSRs plus the extra fraction of FSRs that the sky profile peak is removed from this source. However, in practice, this is not viable as each bin corresponds to a wavelength shift of $10^{-14}$ m, so we would have to know the etalon gap extremely accurately in order to achieve this level of accuracy. It is reasonable to state that this offset value is practically unmeasurable. Therefore a baseline is determined, i.e.,

a sensible estimate of the zero Doppler shift position (see section 3.9.3).

## 3.7 Typical Uncertainties and Signal-to-Noise Ratio

The spectral fitting uncertainties and the signal-to-noise ratio are calculated with the same methods as for the FPI data, which are explained in sections 2.6.2 and 2.6.3. The uncertainties in the fitted sky parameters quantities typically are, for:

- the neutral temperature, ∼90K
- the wind speed, ∼25-30ms$^{-1}$
- the emission intensity, ∼15 counts

and a typical signal-to-noise ratio is ∼3400.

## 3.8 Upgrading to High Resolution

Originally, SCANDI was set up to operate only in the 25 zone configuration. The 51 and 91 zone maps were only at the planning stage. All three of the original zone configurations are displayed in figure 3.3 which shows the sky projections on the left hand side column and the EMCCD chip projections on the right hand side column (courtesy of Yiu, 2012). The 25, 51 and 91 resolution maps are displayed in the top, middle and bottom rows respectively. These have been calculated for a flat area on the sky subject to equation 3.3. However taking the curvature of the Earth into account is found to alter the zonal area by $0 - 27\%$ (Yiu, 2012) for the 25-zone map. This difference is insignificant due to the scales of variability in the thermosphere being relatively large, and so the slight shift in map boundary edges required to implement the curved area projection, is unlikely to have a noticeable impact on the average conditions measured within each zone. Therefore, it is thought unnecessary to implement a curved-area projection. Therefore, the flat (and near-equal) area projection was the first configuration to be implemented, i.e., from 2007-2011.

In 2012, with the onset of solar maximum, higher resolution measurements were possible due to the increased airglow intensity, improving the signal-to-noise ratio at each pixel. SCANDI still required setting up for the two high-resolution zone configurations (51 and 91 zones). An alternative medium resolution zone map of 61 zones (instead of 51) was

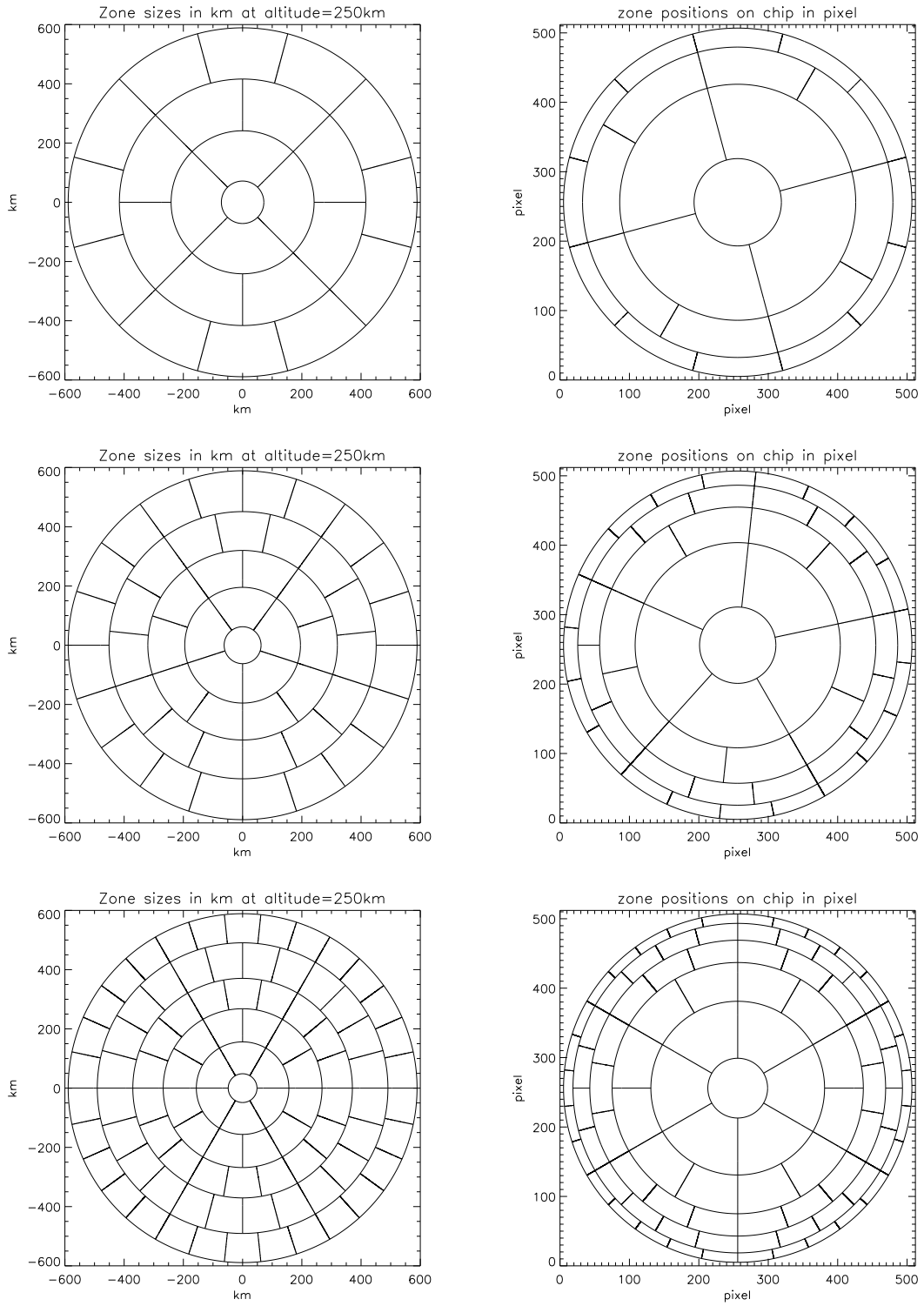


Figure 3.3: SCANDI zone configurations on the sky (left hand column) and on the EMCCD chip (right hand column). The low (25 zones), medium (51 zones) and high (91 zones) zone resolutions are shown in the top, middle and bottom rows (courtesy of Yiu (2011)).

introduced. This map has zones aligned along the geomagnetic meridian, i.e., SCANDI's current North which is where most of the geomagnetic activity is centred, e.g., the cusp (n.b, at times of IMF Bz north, the multi-cell ionospheric convection is not necessarily centred on the geomagnetic meridian). The 51 zone pattern divides the second and forth ring straight through the meridian. Therefore, to retrieve meridian-aligned spectra for these rings would involve averaging zones on either side of the meridian line and losing some resolution. Hence, in usual medium resolution operation, the 61-zone map is typically favoured over the 51-zone map.

In the new maps, the zones were partitioned in the same way as before, so that the projected radial ring separation was constant and the diameter of the central zone was twice this separation. This configuration was chosen so that the zone maps would be easy to interpret.

### 3.8.1 History of SCANDI

It is important to keep track of the different configurations of SCANDI over the years so that we use the data accurately and can combine it correctly within future statistical studies. Figure 3.4 shows a reference table presenting the short history of SCANDI's lifetime up until the season of 2014, summarising all configurations and changes to SCANDI's operation. The first column shows the dates of different periods; the second column shows the filename of the day starting the period; the third column shows the directory in which the data are stored from this period; the fourth column shows if SCANDI is in normal operation, i.e., 'On', or if SCANDI's configuration has been altered; the fifth column shows the clockwise rotation in degrees from geographic north to SCANDI field-of-view north; the sixth column shows the number of zones in the zone-map implemented at the time. The seventh column shows the resolution of the zone maps or the 'Activity parameter', i.e., 'L','M','N' or 'H' which correspond respectively to 25, 51, 61 and 91 zones and is included in the filenames (see column two). This convention was introduced from 19/01/12; before which SCANDI was always in a low resolution 25 zone configuration; not included in the file name but understood. The eighth column shows the 'Alignment parameter' (introduced with the 'Activity parameter' and also included in the filename), i.e., 'M' or 'G' for either magnetically or geographically north aligned fields-of-view; and the ninth column shows the clockwise rotation in degrees from SCANDI's field-of-view north to magnetic north. My work on upgrading SCANDI's zone resolution was carried out between

| Dates dd/mm/yy | Starting File Name (.sdi) | Directory (Wyy/Yyy) | Operation or Changes | SCANDI North (°) | # Zones | Resolution (L/M/N/H) | Alignment (M or G) | $N_{SCANDI}$ to $N_{mag}$ (°) |
|---|---|---|---|---|---|---|---|---|
| 23/12/06 | n/a | n/a | Installation and testing | n/a | n/a | n/a | n/a | n/a |
| 11/02/07 – 04/04/07 | W07211 | W07 | On | -45 | 25 | n/a | n/a | 0 |
| 25/07/07 – 09/08/07 | W07725 | W07 | Testing | -45 | 25 | n/a | n/a | 0 |
| 01/10/07 – 19/11/07 | W07A01 | W07 | On | -45 | 25 | n/a | n/a | 0 |
| 19/11/07 | W07B19 | W07 | Move to KHO | 18.5 | 25 | n/a | n/a | 63.5 |
| 20/11/07 – 02/04/08 | W07B20 | W07 | On | 18.5 | 25 | n/a | n/a | 63.5 |
| 04/08/08 – 29/09/08 | W08804 | W08 | On | 18.5 | 25 | n/a | n/a | 63.5 |
| 30/09/08 | W08930 | W08 | SCANDI sectors re-aligned in software | 18.5 | 25 | n/a | n/a | 63.5 |
| 01/10/08 – 21/11/08 | W08A01 | W08 | On | 18.5 | 25 | n/a | n/a | 63.5 |
| 22/11/08 | W08B22 | W08 | SCANDI North re-aligned in software | 18.5/-32 | 25 | n/a | n/a | 63.5/0 |
| 23/11/08 – 03/04/09 | W08B23 | W08 | On | -32 | 25 | n/a | n/a | 0 |
| 18/09/09 – 08/01/10 | W09918 | W09 | On | -32 | 25 | n/a | n/a | 0 |
| 06/11/10 – 16/11/10 | Y10B06 | Y10 | On | -32 | 25 | n/a | n/a | 0 |
| 18/11/10 – 14/02/11 | W10B18 | W10 | On | -32 | 25 | n/a | n/a | 0 |
| 15/02/11 - 20/03/11 | Y11215 | Y10 | On | -32 | 25 | n/a | n/a | 0 |
| 20/10/11 – 06/12/11 | Y11A20 | Y11 | On | -32 | 25 | n/a | n/a | 0 |
| 07/12/11 – 18/01/12 | n/a | n/a | Testing new zone configurations | -32 | n/a | n/a | n/a | 0 |
| 19/01/12 – 03/04/12 | WNM12119 | W11 | On | -32 | 61 | N | M | 0 |
| 25/10/12 – 12/11/12 | WNM12A25 | W12 | On | -32 | 61 | N | M | 0 |
| 13/11/12 – 29/04/13 | WHM12B13 | W12 | On | -32 | 91 | H | M | 0 |
| 11/11/13 – 19/03/14 | WNM13B11 | W13 | On | -32 | 61 | N | M | 0 |
| 10/10/14 – 04/04/15 | WNM14A10 | W14 | On | -32 | 61 | N | M | 0 |

Figure 3.4: A table presenting a history of SCANDI's lifetime up to season of 2014. The columns show i) dates of different periods; ii) shows the filename of the day starting the period; iii) the data-storing directory from this period; iv) if SCANDI is in normal operation ('On') or if the configuration has been altered; v) the clockwise rotation ($^o$) from geographic north to SCANDI's FOV north; vi) the number of zones in the map implemented during the period; vii) the resolution of the zone maps or 'Activity parameter', i.e., 'L', 'M', 'N' or 'H' for 25, 51, 61 and 91 zones (included in filename from 19/01/12, see column two); viii) the 'Alignment parameter' (introduced with the 'Activity parameter' and also included in the filename), i.e., 'M' or 'G' for magnetically or geographically north aligned FOV; and viiii) the clockwise rotation ($^o$) from SCANDI's FOV north to magnetic north.

07/12/11 - 18/01/12 in the 'Testing new zone configurations' period. From 19/01/12 the 51, 61 and 91 zone equal-radially separated rings sky projection configurations were installed in SCANDI's control software and have been in operation ever since.

## 3.9 Sky Parameter Calibration

After data reduction some calibrations are required on the data before data analysis is performed. There is an emission intensity offset that needs to be corrected and a baseline needs to be determined in order to estimate the absolute LOS winds. It is difficult to calibrate the neutral temperature because there is little data to cross-compare with. The only measurements of the thermosphere region of our emissions (120km and 240km altitude), separate to the FPIs, are taken in situ by rockets. However, these campaigns

are too infrequent and the datasets do not have the temporal continuity to be suitable for deriving a reliable calibration function. Therefore, the temperatures are left uncalibrated.

### 3.9.1 Emission Intensity

Figure 3.5 shows the averaged 630nm intensities measured by SCANDI from 2012 to 2015, from November to January. The top plot displays them as a sky-projection zone map of SCANDI's field-of-view (FOV), whereby the different colour zones represent different levels of intensity (defined within the colour-bar underneath). Plots formatted like this are from here on termed map plots. The bottom plot displays the average intensity with zenith angle. The map plot clearly shows a systematic radial intensity variation centred around the zenith zone (i.e., the central zone) of SCANDI's FOV which is also shown in the dramatic reduction in intensity with zenith angle in the line plot below. Possibilities that this effect is a real 630nm source, another atmospheric phenomenon or an instrumental artefact are considered.

It is unlikely that this radial intensity gradient (or ring offset) is a real localised 630nm source centred about SCANDI's zenith zone in the thermosphere, even though SCANDI is intentionally positioned to drift through the cusp. The cusp location is variable and travels from magnetic east to west across the FOV so would create more of a swathe of enhanced brightness. Furthermore, the upper atmosphere is generally understood to be too dynamic a medium to create such a perfectly concentric ring-offset. Also there is no such ordering in the corresponding temperature (see figure 3.10) or wind data (see figure 3.11). Some correspondence would be expected as this represents intense auroral activity, which if present, would enhance temperatures and influence winds. Furthermore, it cannot be an instrumental effect as the calibration lamp intensities do not exhibit this ring offset. There will still be some less-ordered signature of the cusp within this map plot which is likely to enhance the ring-offset. As mentioned above it will be less concentric and more swathe-shaped. When determining the correction factors for each ring, care should be taken not to over-correct and remove the cusp signature. It seems probable that the phenomenon is atmospheric in origin and a result of different optical paths associated with different zenith angles. Two known zenith-dependent mechanisms which affect incident intensity from an extended light source are:

- atmospheric extinction, i.e., scattering and absorption by atmospheric molecular

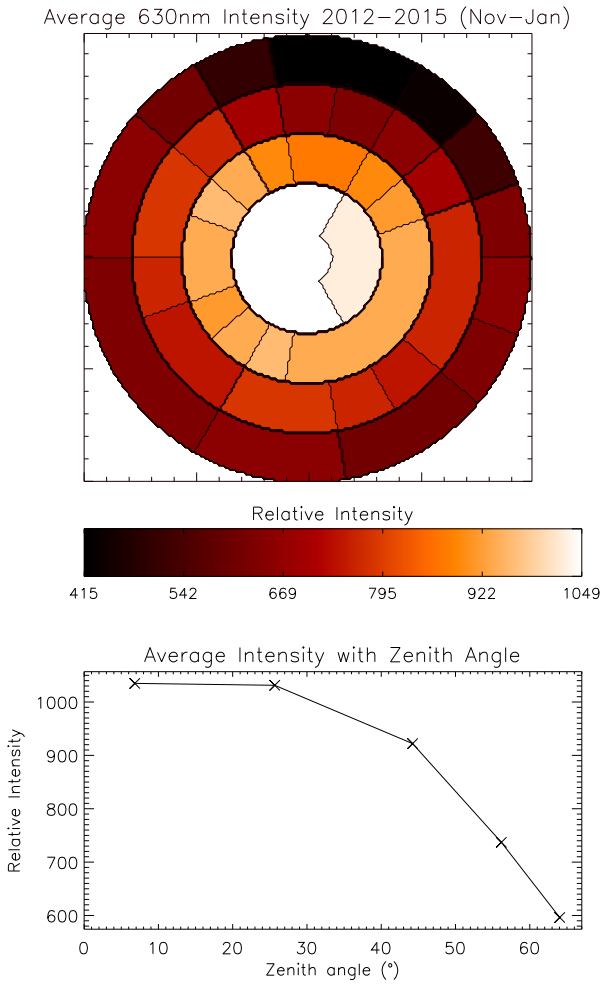


Figure 3.5: Averaged 630nm intensities measured by SCANDI from 2012 to 2015 (November-January inclusive). The top plot displays them as a sky-projection zone map of SCANDI's field-of-view (map-plot), whereby the different colour zones represent different levels of intensity (defined within the colour-bar underneath). The bottom plot is the average intensity with zenith angle.

species and aerosols

- Van Rhijn effect, i.e., slantwise path of incident light means larger number of emitting atoms from a uniform emitting layer

These two effects are explained and addressed in the following subsections.

#### Atmospheric Extinction

There is increased extinction for light received from a larger zenith angle whereby it travels a longer optical path and so the incoming light will have undergone more scattering and

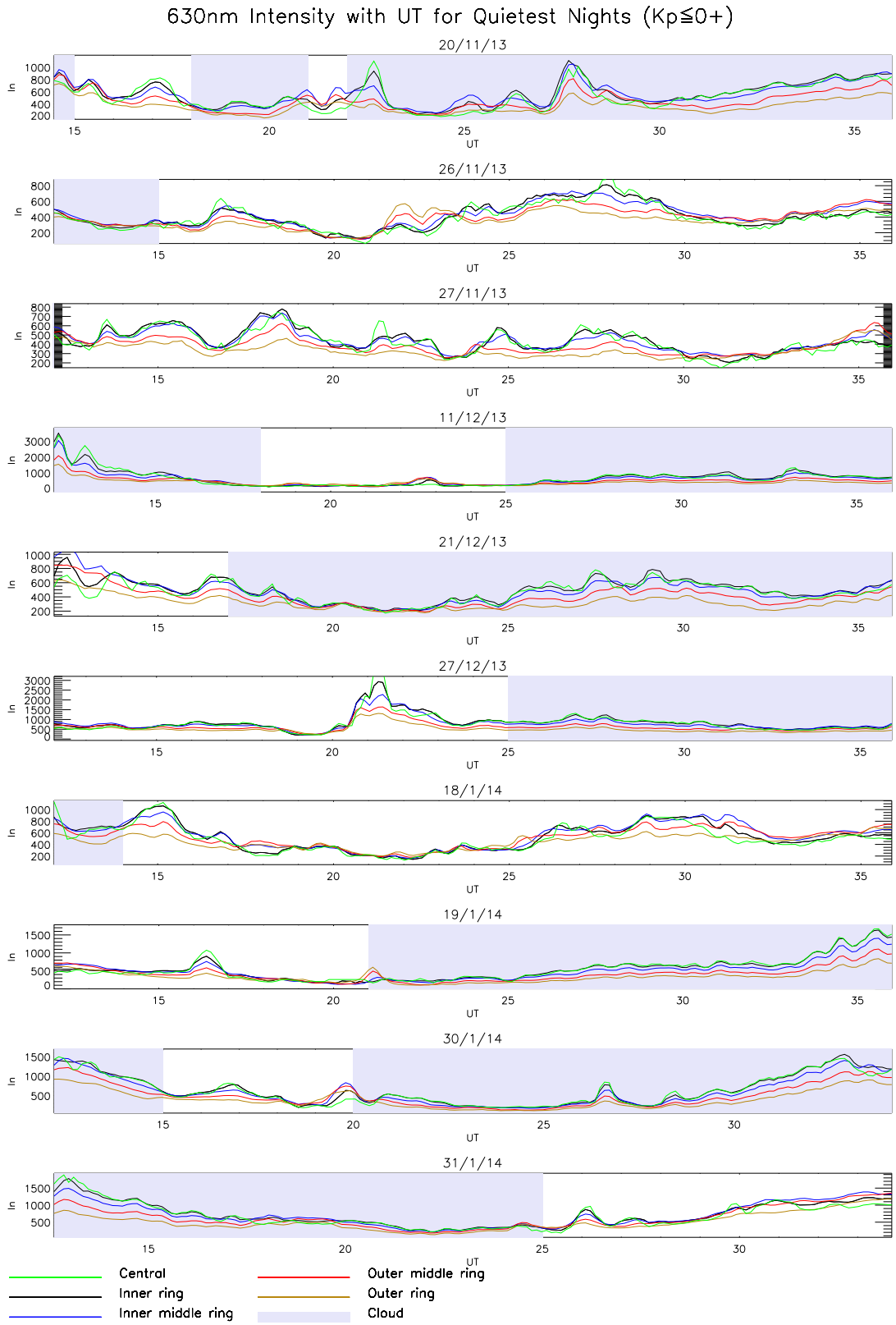


Figure 3.6: Ring-averaged 630nm intensity line plots with UT for quietest nights in dataset ($K_P < 0+$). From zenith to inner to outer ring colour-coded green, black, blue, red and gold. Cloudy periods shaded in lavender.

absorption. Therefore the light from the outer rings of SCANDI appears dimmer than the inner ones. This effect is described by the Beer-Lambert law:

$$I = I_0 \exp(-m\tau) \tag{3.4}$$

where $I_0$ is the zenith intensity, $\tau$ is the optical depth along the observation line-of-sight zenith angle, $\theta$, m is the optical air mass which, for a plane-parallel atmosphere, is equal to $\sec(\theta)$. The optical depth is a measure of the attenuation of a transmitted radiative power. By definition it is the logarithm of the ratio between incident to transmitted power, hence it is a dimensionless number. The quantity increases with optical path length and may also be defined as:

$$\tau = \int_0^l \alpha(z)dz \tag{3.5}$$

where l is the optical path length and $\alpha(z)$ is the attenuation coefficient dependent on the depth along the optical path, z, which is along the line-of-sight observation zenith angle, $\theta$. If assuming a linear attenuation, i.e., $\alpha(z)$ is uniform along the optical path, then the optical depth becomes $\tau = \alpha l$.

Following is an attempt to model the optical depth using the Beer Lambert law. The method is to plot the logarithm of $I/I_0$ against $\sec(\phi)$ and to determine the optical depth from the gradient, $-\tau$. Firstly, a sensible dataset needs to be chosen which encodes the atmospheric extinction effect but has no dominant auroral emissions which will contaminate our corrections. The SCANDI dataset is limited to November to January data as these are the only months with corresponding cloud record files. The 61-zone configuration SCANDI data are used which limits the dataset to 2012 to 2015. Out of this dataset 11 days were found with $K_P \leq 0+$ representing extremely geomagnetically quiet conditions, where auroral arcs are most unlikely to be present. This is the chosen dataset for this study of intensities. These nights' intensity variations are plotted with UT in figure 3.6. The intensity values are ring-averaged and are colour coded from zenith zone to inner ring to outer ring, green, black, blue, red and gold. The cloud periods are shaded in lavender. Throughout all nights the ring offset is clearly evident however it seems more pronounced at cloudy times compared with the clear periods. This is likely to be due to the clouds removing the directional information from the photons and so effectively creating a homogeneous intensity light source. Therefore the differentiation in intensity between the rings

can only be due to the extinction from the base of the clouds to the ground. Even though an accurate recording of the extinction effect, it fails to record the additional extinction that occurs from the thermosphere to the cloud altitude in the troposphere. However, the majority of the extinction is expected to occur in the troposphere where three-quarters of the mass of the whole atmosphere lies. Nonetheless, the clear data are still a worthy record of the full emission-to-ground extinction effect as there are no intense auroral structures present (so the sky variations in intensity should be minimal). As it is unclear whether the clear or cloudy dataset is more reliable, both are used to model the optical depth.

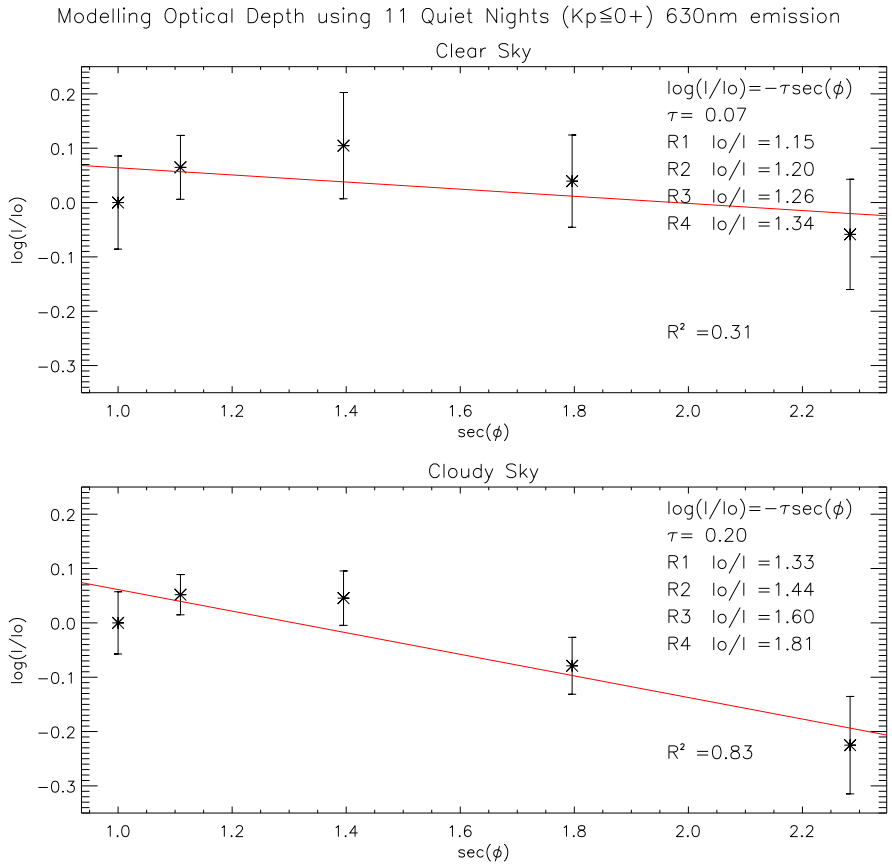


Figure 3.7: Plot of the logarithm of intensity normalised to the zenith intensity against the secant of the zenith angle for a uniform intensity period from the night of the 21st January 2012. The Beer Lambert law shows the gradient to be the optical depth.

$I_0$ is taken to be the zenith intensity as the absolute transmitted intensity is unknown. The zenith angle, $\phi$, is taken as the zenith intersecting the centre of each ring. Figure 3.7 shows plots of the logarithm of $I/I_0$ against sec($\phi$) (where $I$ is the intensity for each ring) for both the clear sky data (top) and the cloudy sky data (bottom). The data has been binned by zenith angle $\phi$ and averages for each bin are taken. The error bars

represent the standard deviation for each bin. Both plots have a least-squares fitted line overplotted (red) with a corresponding goodness of fit $R^2$ value calculated and labelled on the bottom right of the plots. The gradient estimates an optical depth value of $\tau = 0.06$ for the clear data and $\tau = 0.19$ for the cloudy sky data. Substituting these values back into the Beer-Lambert equation with the zenith angles and re-arranging, we calculate the correction factors $I_0/I$ which is listed in the top right of each plot. These can then be used to normalise each rings' intensities to the zenith zone level.

The clear sky plot shows a weak linear relationship ($R^2$=0.31) between $\log(\frac{I}{I_0})$ and $\sec(\phi)$ whereas the cloudy sky plot shows a strong linear relationship ($R^2$=0.83). Therefore, evidently the clear sky data does not detect a Beer-Lambert law relationship whereas the cloudy sky does. The $R^2$ value for the cloudy (clear) sky data of 0.83 (0.31) means, if given some zenith angles of observation ($\phi$), the intensity values calculated using the relationship of this fitted line, could correctly predict 83% (31%) of the variablity of the actual intensity values. Therefore the clear sky data's relationship offers a poor correction method to SCANDI's ring offset, however, the cloudy sky data's is reasonable. Nevertheless, there are other methods to explore that could more accurately correct the ring offset, e.g, the Van Rhijn effect and the Herrero and Meriwether methods (see sections 3.9.1 and 3.9.1).

**Van Rhijn Effect**

The Van Rhijn effect (Chamberlain et al, 1961) states that light received from larger zenith angles from a uniform-thickness emitting volume appear brighter due to the fact that the light from this layer has a longer path through the emitting layer and so more atoms are emitting light. Chamberlain et al (1961) provide a formula which calculates the correction factors for different zeniths $I(\theta) = V(z|\theta)I_0$ where

$$V(z|\theta) = \left(1 - \frac{a^2}{(a+z)^2}\sin^2\theta\right)^{-1/2} \tag{3.6}$$

where a is the horizontal distance from the observing volume to the location directly overhead the observing site at the altitude of the emission. The correction factors derived from the Van Rhijn effect, the set [1.00, 0.83, 0.67,0.55, 0.47] (from zenith to outer ring), will clearly have a diminishing effect on the larger zenith intensities which is the opposite

effect desired (the aim of correcting the ring-offset is to amplify the larger zenith intensities to produce a constant detected intensity level across the FOV). If the Van Rhijn effect is occurring in the atmosphere, it is clearly not powerful enough to be detected by SCANDI, therefore this effect is not investigated any further.

**Herrero and Meriwether Correction Method**

The shortcomings of the attempts to model the extinction function theoretically (Beer-Lambert and Van Rhijn) could be due to various factors:

- the Beer-Lambert law is too idealised and the atmosphere's behaviour does not follow the relationship exactly.
- uncertainty in the emission layer height (assumed to be at 250km altitude) meaning the $\sec(\theta)$ values are not representative of the same altitude emission or region of intensity enhancement, adding noise to the measurements.
- inhomogeneous sensitivity across the EMCCD detector biasing the intensity measurements across the FOV.

Herrero and Meriwether (1980) present a more direct approach to determining an intensity correction factor calculated empirically. It involves using quiet night data where the intensities are reasonably symmetric about the zenith. The intensity signals are three-point-smoothed and normalised to the zenith intensity. The average percentage difference between each ring is calculated and assumed to represent the Van Rhijn intensity distribution, convolved with a function that represents the atmospheric extinction. However, the function will also incorporate any other deviations from theoretical behaviour, such as listed in the above items.

The dataset used (to calculate the correction factors) is the quiet-night data displayed in figure 3.6 which clearly displays the intensities being reasonably symmetric about the zenith as required. Correction factors are found in three different ways: from using, i) the clear sky data, ii) the cloudy sky data and, iii) all the data together (both clear and cloudy). All the data together are checked as this procedure incorporates the advantages from both the clear and cloudy-sky data mentioned previously. Initially these different methods are tested on one night of data. The night chosen is the 21st January 2012 (later presented in this thesis as a case study, the CUSPN campaign in chapter 8. Figure 3.8

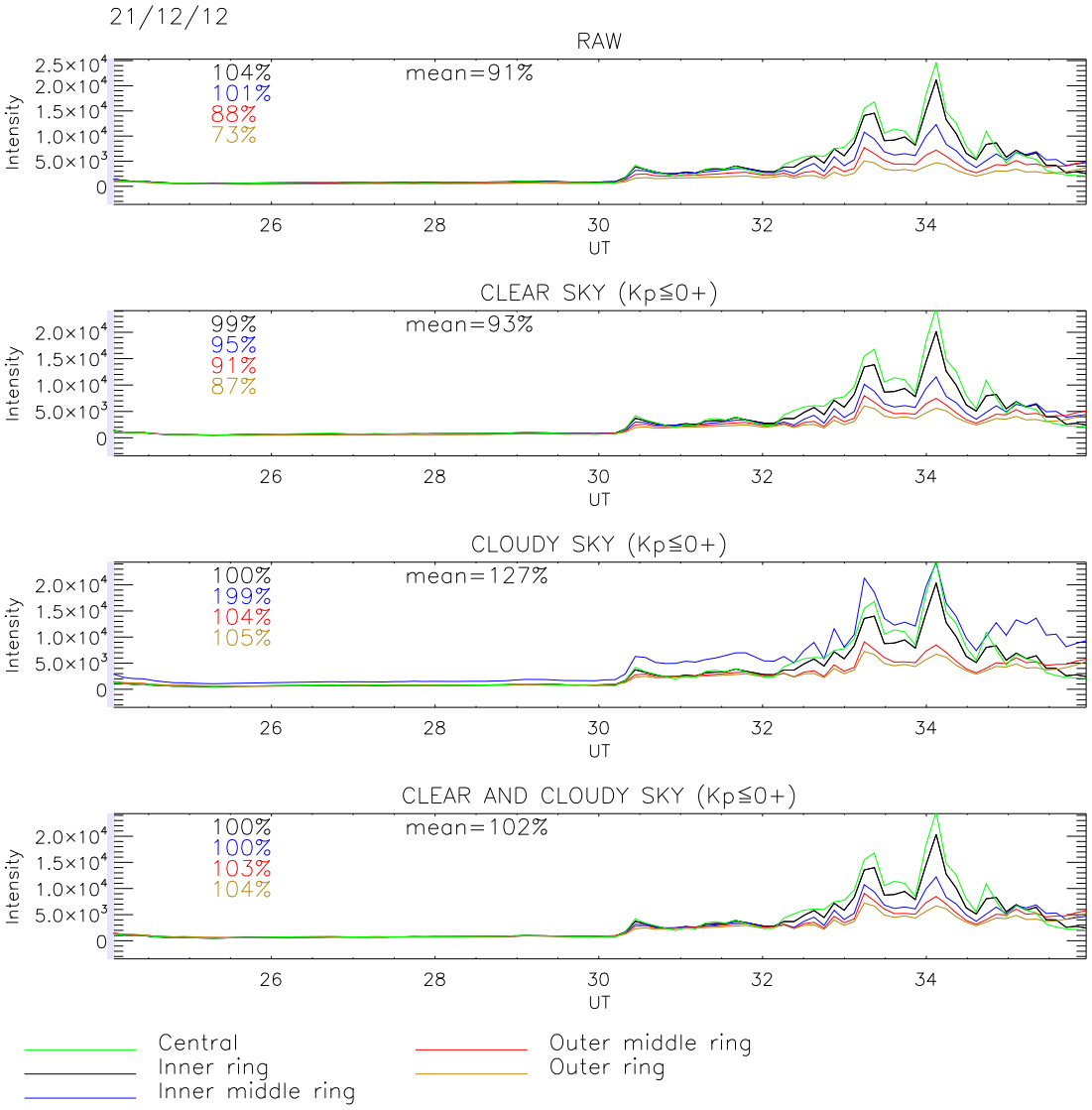


Figure 3.8: Ring-averaged 630nm emission intensity with UT for 21/01/12 corrected by different methods. The top plot shows the raw uncorrected data, the next plot shows the data corrected by the average difference in ring intensity from the quietest of SCANDI's nights ($K_P < 0+$) using i) clear sky data, ii) cloudy sky data and iii) all sky data. The percent difference for each ring compared to the zenith is listed in the top left of each plot with the mean of these labelled to the right.

shows this data as ring-averaged intensities as a function of UT, corrected by different methods. The top plot shows the raw uncorrected data, the next plot shows the clear sky data, the next the cloudy sky data and the last, both clear and cloudy sky data. The percentages listed on the top left of each plot show the percent difference for each ring's average intensity compared to the zenith intensity (colour-coded for each ring). The mean of these, is labelled to the right.

From the all-sky camera the case-study night (of figure 3.8) shows no visible aurora until 30UT (i.e., 06UT the following day), however during the following interval, there is a relatively high level of aurora creating large intensity gradients across the field of view. This increased emission manifests as a large ring offset which is evident in the plot from 30-36UT (i.e., 06-12UT) and peaks at 34UT (i.e., 10UT). This is not, however, the effect for which we wish to correct for. Our aim is to correct the ambient atmosphere ring offset occurring before 30UT, which is not clear in this plot. This is the reason geomagnetically active nights are not included in deriving the correction factors; the auroral features would create falsely enhanced ring offsets and so the correction factors derived would overcompensate. For this reason, the percent differences listed on the plots are calculated with the data from 24-30UT (i.e., 00-06UT) when the sky is clear and geomagnetic activity is low $K_P$=[1-,1+].

The raw uncorrected data shown in the top plot of figure 3.8 show an average intensity offset between the rings and the zenith of 91%; with the separate rings' differences with the zenith intensity ranging from 73-104%. From the remaining plots in figure 3.8, we want to select the correction method (based on different sky-condition data), which shows an average percent difference in intensity (between all rings) as close to 100% as possible. 100% represents the zenith zone intensity and so if the other rings are also at (or close to) 100% this means that there is no intensity gradient and the ring-offset is successfully removed. Evident on the figure, the optimum correction method is the one which derives its correction factors from using both clear and cloudy sky data (fourth plot). This gives an average percent difference (between the rings and the zenith intensity) of 102% (an 11% increase on the raw uncorrected data's average offset in the first plot (91%)). The separate ring's differences range from 100-104%; a strong match to the zenith intensity. This is a good indicator that this is the optimum correction method. However we need a larger statistical sample to confirm that this is the case.

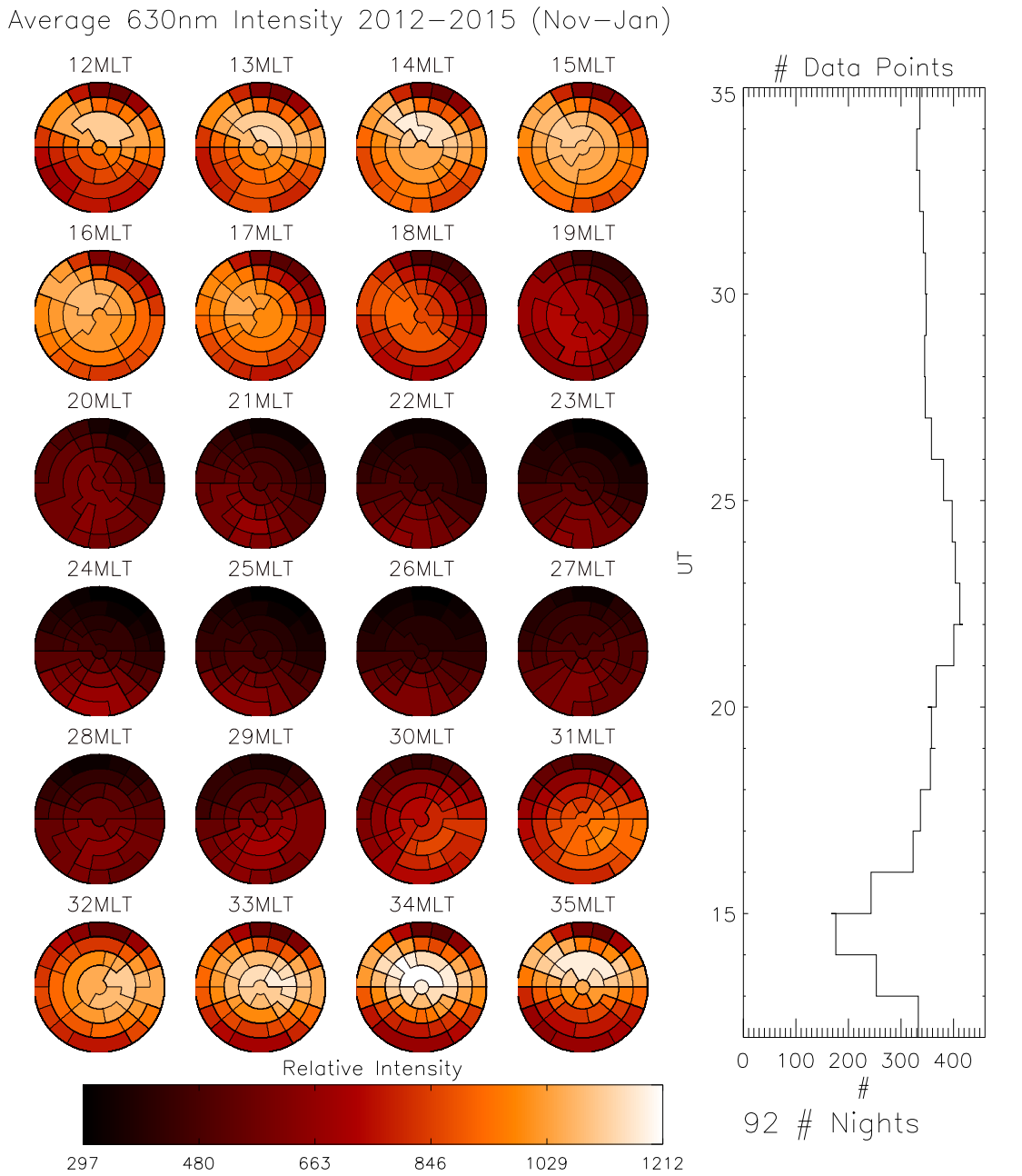


Figure 3.9: Average 630nm intensity from 2012-2015 (November to January, clear sky only) displayed as hourly UT map-plots in a map plot format.

Figure 3.9 shows all clear-sky SCANDI 630nm intensities from November to January 2012-2015 displayed as map plots with magnetic local time (MLT). The data are corrected by the Herrero and Meriwether (1980) method including both clear and cloudy sky data to determine the correction factors.

92 days are included in the averages and the remaining sample statistics are displayed in the side histogram to the right of the figure. The number of data points for each hourly average range from ∼180-410. All geomagnetic activities are included meaning that auroral features such as the cusp should show up. The data are split by MLT so that it is possible to distinguish the auroral features from a poorly corrected ring offset. We are expecting to see the cusp about 12MLT which might appear as a swathe of brightness across the FOV. At 12MLT its signature is evident with peak intensities slightly to the north of zenith in a swathe shape. At 31MLT the soft precipitation is seen entering the field-of-view from the east and at 18MLT it can be seen leaving the field-of-view from the west. Between these times when it seems the cusp is overhead, there will be a true ring offset. However outside of these times, between 19-29UT (i.e., 19-05UT) no ring offset should be present if the correction is applied correctly. Clearly on these map plots there is a smooth gradient across the field of view with no ring offset. This is sufficient proof that the Herrero and Meriwether (1980) method including both clear and cloudy sky data comprises a good correction method for present purposes. This algorithm will be used routinely herein to correct the intensity data.

### 3.9.2 Temperature

The neutral temperature can produce a ring offset when the etalon plates are not completely parallel. Theoretically the face of an etalon can be divided into an infinite number of 'mini-etalons'. Each mini-etalon will produce an interference pattern of concentric rings (from light passing through the optics). The resultant interference pattern is the superposition of each 'mini-etalon's' ring pattern. If the etalon is completely parallel then all the rings will superimpose exactly. If not, then the different spacings between the plates at each point will give rise to larger and smaller radii rings, which when superimposed, give a broader resultant peak. This phenomenon is explained in depth in Yiu (2012). The non-parallelism of the etalon is due to drift of the analogue electronics (used for the scanning) not maintaining a zero differential gain between the three channels associated with the three legs of the etalon. The plates were parallelised and fine-tuned by Dr Ian

McWhirter in December 2009 and, since then, there has been no offset. As we may see from the average neutral temperature map plot (figure 3.10), there is no ring offset and therefore, no further calibration is required.

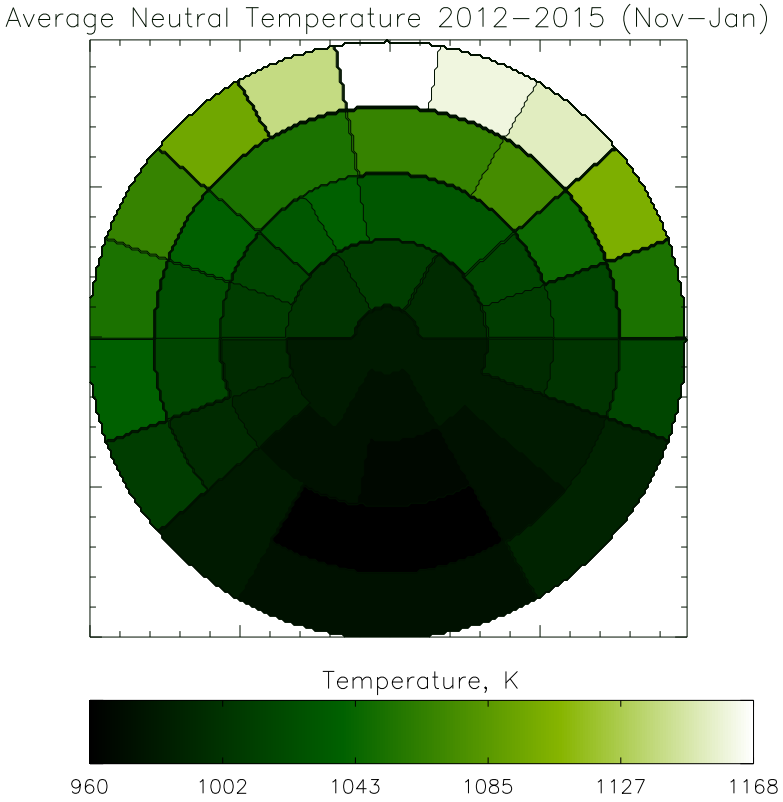


Figure 3.10: Map plot of SCANDI's raw neutral temperature derived from 630nm emission averaged over 2012-2015 (November to January).

### 3.9.3 LOS Wind

The main calibration process for SCANDI's LOS is determining its zero velocity baseline. The method that is employed for the SCANDI dataset uses a zonally dependent average as a baseline (Yiu, 2012). There has always been some uncertainty on the origin of the baseline to be removed. This issue is investigated within this section.

SCANDI's wind measurements are relative and a zero-velocity reference needs to be determined in order to estimate their absolute value. Figure 3.11 displays SCANDI's raw line-of-sight (LOS) wind measurements averaged over 2012-2015 (November to January) as a map plot. The colour-bar underneath the plot shows the range of winds as -179ms$^{-1}$ to -70ms$^{-1}$; all negative, i.e., wind velocity towards the observer. After the baseline is

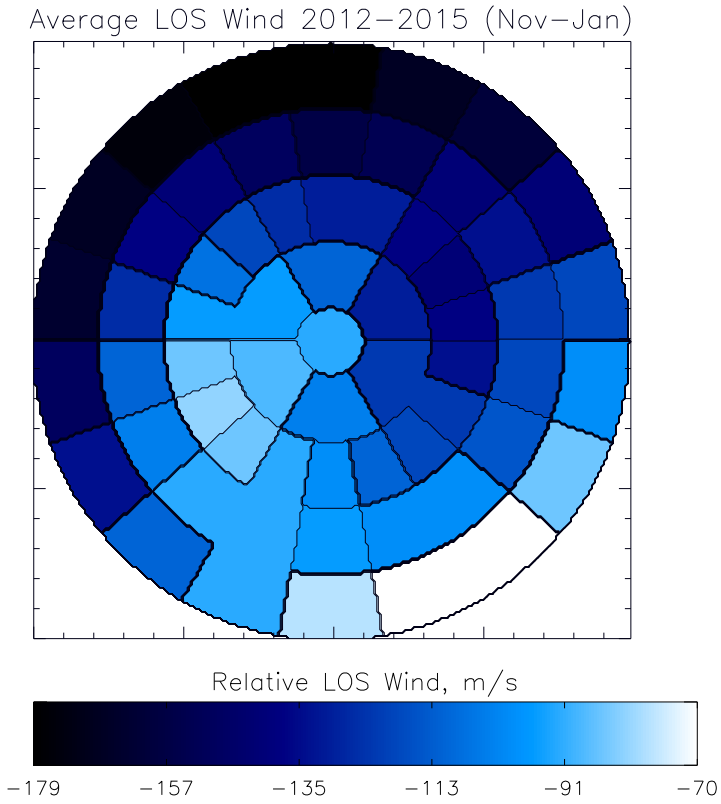


Figure 3.11: Map plot of SCANDI's relative line-of-sight winds averaged over 2012-2015.

removed some of these values will be positive, i.e., wind velocity away from the observer. This negative value is imposed on the wind measurements within the phase mapping process (see section 3.4). The negative shift represents the bin position of the laser light which is used as a zero-velocity reference for all zones. This shift value, i.e., the baseline, needs to determined in order to re-centre the data, so that the less negative values in the range become the positive values. It cannot be determined theoretically as the laser peak lies several orders of interference away from the sky peak and there are too many bins between them for a precise value to be estimated; a 1-bin error would result in a significant error in the calculated Doppler shift, i.e., $42\text{ms}^{-1}$ compared to typical error values of $\sim 25\text{ms}^{-1}$. The baseline needs to be estimated empirically from the derived data. Figure 3.11 also shows a LOS wind field negative to the north east and less negative to the south west. A consideration needs to be given to this either being a real background wind field or an instrumental effect.

Currently the baseline employed is a zonally dependent average of the season's November to January data. Yiu (2012) found this to fit with the FPI data best in correlation

and absolute value and this is the method used to correct the data throughout this thesis.

## 3.10 Vertical Wind Assumption

There are two ways to treat the vertical wind component present in SCANDI's off-zenith line of sight wind measurements. Either it can be neglected, because the vertical winds are typically so much smaller in magnitude than the horizontal winds, or, it can be assumed to be constant across the whole field of view and equal to the zenith vertical wind and then subtracted from all zones.

In the former approach, the assumption of a zero vertical wind field is reasonable in that the typical vertical wind is of the order of tens of $ms^{-1}$ and the horizontal wind of the order hundreds of $ms^{-1}$. With the line of sight component taken, this amount is even smaller and will have a negligible effect on the line of sight winds, possibly even within their error bars. However the atmosphere is frequently not so well behaved and several observations have been made of vertical winds exceeding 100 $ms^{-1}$ (Spencer et al, 1982; Crickmore et al, 1991; Innis et al, 1996 ,1999; and Price et al, 1995). If these are true large-scale vertical wind events, correcting for them will introduce large unwanted errors. However, there is a chance that these authors may be mistaking a hydroxyl contamination for a true vertical wind as has been studied in later in this thesis in chapter 5. A more sensible upper bound of the vertical wind magnitude is found to be $\sim$70$ms^{-1}$ (see left tail of histogram in figure 4.9 in chapter 4).

In the latter approach, using the zenith vertical wind, the assumption of a uniform vertical wind field across the entire field of view is a precarious one, due to the vertical wind's highly varied and unpredictable horizontal scale size (Crickmore, 1993, Kosch et al, 2000; Ishii et al, 2001, Anderson et al, 2011). To recap, it is found that the horizontal scale size is often $\sim$200km; however, it can range from less than $\sim$50km to a value as large as 800km (see section 1.4.2). This means that, due to SCANDI's large field of view (1076km across) the zenith vertical wind will very rarely be a reasonable approximation for that across the whole field of view. This means that, when correcting for a vertical wind using the zenith only, the inner rings of SCANDI will be corrected for accurately and moving radially outwards over the field of view, at some point the unknown horizontal vertical wind scale size will be crossed. Beyond this point, the correction will just introduce additional error. This argument can also be extended to the former approach (assuming

a negligible vertical wind field). If the vertical wind were found to be negligible at the zenith zone, it cannot be reliably assumed that the same negligible vertical wind will be experienced 500km away in the outer rings.

## 3.11 Horizontal Wind Fit

It is possible to fit the baseline-corrected line of sight winds to a two-dimensional horizontal wind vector field. The method for determining the horizontal wind fields from LOS measurements originated from Browning and Wexler (1968), who applied the technique to azimuth-scanned Doppler radar measurements. Burnside et al (1981) adapted this algorithm to analyse thermospheric airglow observations. Finally Conde and Smith (1998) further adapted it into the method that is derived following. The geometric layout and FOV coordinate system is specific to SCANDI. Therefore, the trigonometry in the algorithm following is slightly different to other records of this algorithm (Conde and Smith, 1998; Anderson, 2012).

Clearly, it is impossible to determine the horizontal wind vector from the line of sight unambiguously so two assumptions are necessary:

- the vertical wind is constant across the entire FOV
- the vertical winds are small compared to the typical horizontal wind of hundreds of $ms^{-1}$

SCANDI observes spectra from several rings centred around the zenith with different radii specified by their zenith angle, $\phi$ (n.b., change of symbol for zenith angle from $\theta \rightarrow \phi$ as is the notation in Conde and Smith (1998)). Each ring is divided into n equal sized zones with the $k^{th}$ zone centred on azimuthal angle $\theta_k$. The set of LOS wind speeds in a ring is denoted as, $\{V^{||}(\theta_k, \phi)\}$, the vertical component is denoted as $\{Z^{||}(\theta_k, \phi)\}$ and the horizontal as $\{H^{||}(\theta_k, \phi)\}$. Thus:

$$V^{||}(\theta_k, \phi) = Z^{||}(\theta_k, \phi)cos(\phi) + H^{||}(\theta_k, \phi)sin\phi \tag{3.7}$$

The first assumption is applied where a uniform vertical wind, $V_z$, can be related to a LOS wind as:

$$Z^{||}(\theta_k, \phi) = V^{||}(\theta_0, 0) = V_z \tag{3.8}$$

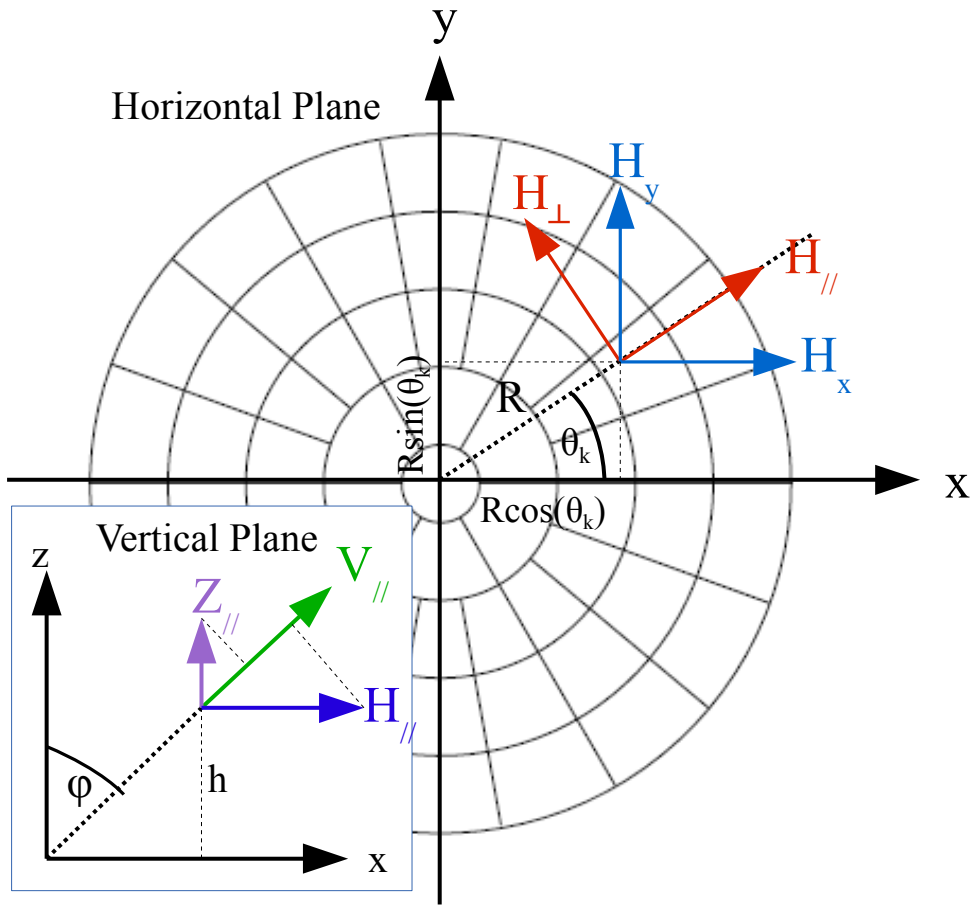


Figure 3.12: Angles and vectors relevant to SCANDI's trigonometric set-up.

This assumption allows the vertical wind a contribution to be removed from the measured LOS wind from each zone. Therefore the LOS wind that only contains the horizontal component is:

$$H^{||}(\theta_k, \phi) = \frac{V^{||}(\theta_k, \phi) - V_z \cos(\phi)}{sin\phi} \tag{3.9}$$

In the horizontal xy plane, let $\theta_k$ be the anticlockwise angle from the positive x axis around the origin, i.e., the horizontal location of the observatory site. It should be noted that this coordinate system is different to Conde and Smith (1998) who use a clockwise bearing from the y axis. Therefore, the trigonometry calculated as follows is specific to the SCANDI SDI zone configuration. Subject to our reference frame, $H^{||}(\theta_k, \phi)$ can be described in terms of the spatially varying zonal, $H_x$, and meridional, $H_y$, wind fields:

$$H^{||}(\theta_k, \phi) = H_x \cos(\theta_k) + H_y \sin(\theta_k) \tag{3.10}$$

$H_x$ and $H_y$ can then be expanded as a first order Taylor series about the zenith:

$$H_x = u_0 + \frac{\partial u}{\partial x}x + \frac{\partial u}{\partial y}y \tag{3.11}$$

$$H_y = v_0 + \frac{\partial v}{\partial x}x + \frac{\partial v}{\partial y}y \tag{3.12}$$

where x and y are the zonal and meridional distances from the zenith to a zone, at zenith angle, $\phi$, and azimuth angle, $\theta_k$. For an observed emission altitude at mean height h, $x(\theta_k, \phi)$ and $y(\theta_k, \phi)$ can be given in terms of radial distance, $R = h\tan(\theta_k)$:

$$x(\theta_k, \phi) = h\tan(\phi)\cos(\theta_k) = R\cos(\theta_k) \tag{3.13}$$

$$y(\theta_k, \phi) = h\tan(\phi)\sin(\theta_k) = R\sin(\theta_k) \tag{3.14}$$

Combining equations 3.10-3.12 gives:

$$\begin{aligned} H^{||}(\theta_k, \phi) = u_0\cos(\theta_k) + (\frac{\partial u}{\partial x})R\cos^2(\theta_k) + (\frac{\partial u}{\partial y})Rsin(\theta_k)\cos(\theta_k) \\ +v_0\sin(\theta_k) + (\frac{\partial v}{\partial y})R\sin^2(\theta_k) + (\frac{\partial v}{\partial x})Rsin(\theta_k)cos(\theta_k) \end{aligned} \tag{3.15}$$

Substituting in the trigonometric identities:
$\sin(x)\cos(x) = \frac{1}{2}\sin(2x)$
$\sin^2(x) = \frac{1}{2} - \frac{1}{2}\cos(2x)$
$\cos^2(x) = \frac{1}{2} + \frac{1}{2}\cos(2x)$
into equation 3.15 gives:

$$\begin{aligned} H^{||}(\theta_k, \phi) = u_0\cos(\theta_k) + (\frac{R}{2})(\frac{\partial u}{\partial x})(1 + cos(2\theta_k)) + (\frac{R}{2})(\frac{\partial u}{\partial y})sin(2\theta_k) \\ +v_0\sin(\theta_k) + (\frac{R}{2})(\frac{\partial v}{\partial y})(1 - \cos(2\theta_k)) + (\frac{R}{2})(\frac{\partial v}{\partial x})\sin(2\theta_k) \end{aligned} \tag{3.16}$$

For a given zenith angle $\phi$, the wind field is sampled at n independent azimuths (or zones in a ring), the horizontal component of the LOS wind can also be represented as a Fourier series by:

$$H^{||}(\theta_k, \phi) = a_0 + \sum_{m=1}^{\frac{n}{2}-1}(a_m\cos(m\theta_k) + b_m\sin(m\theta_k)) \tag{3.17}$$

which can be inverted to give the Fourier coefficients as follows:

$$a_0(\phi) = \frac{1}{n}\sum_{k=0}^{n-1} H^{||}(\theta_k, \phi) \tag{3.18}$$

$$a_m(\phi) = \frac{2}{n}\sum_{k=0}^{n-1} H^{||}(\theta_k, \phi)\cos(m\theta_k) \tag{3.19}$$

$$b_m(\phi) = \frac{2}{n}\sum_{k=0}^{n-1} H^{||}(\theta_k, \phi)\sin(m\theta_k) \tag{3.20}$$

The Fourier coefficients can be compared with the explicitly written out Taylor expansions to determine the unknown terms, $u_0$, $v_0$ and the partial derivatives. Comparing coefficients for all the sinusoidal terms gives:

$$u_0 = a_1 \tag{3.21}$$

$$v_0 = b_1 \tag{3.22}$$

$$a_0 = \frac{R}{2}(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) \tag{3.23}$$

$$a_2 = \frac{R}{2}(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}) \tag{3.24}$$

$$b_2 = \frac{R}{2}(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}) \tag{3.25}$$

The values of the Fourier coefficients are calculated directly from the measured horizontal wind data, $H_{||}(\theta_k, \phi)$ and therefore the background wind terms, $u_0$ and $v_0$, can be resolved immediately with equations 3.21 and 3.22. With four or more data points in one annulus the background wind terms can be calculated. However if there are six or more data points then the partial derivatives may also be found. The former scenario only arises for the inner ring of the low resolution 25 zone sky map. In this case, we assign an average of the other rings' partial derivatives whilst keeping the ring's own background wind terms. In all other rings for all other resolution sky maps there are 6 data points and so further calculation is required. There are three remaining equations to be solved: $a_0$, $a_2$ and $b_2$, with four unknowns, $\frac{\partial u}{\partial x}$, $\frac{\partial v}{\partial x}$, $\frac{\partial u}{\partial y}$ and $\frac{\partial v}{\partial y}$. Hence, there are insufficient terms to uniquely calculate all these partial derivatives therefore an assumption needs to be made for one of them. It is sensible to make the assumption for the $\frac{\partial v}{\partial x}$ term as this shows

the least spatial variation. At auroral latitudes, the zonal wind field is likely to be the most variable. This is due to it being subject to the ion drag driven by the convection electric field which is known to have very high temporal and spatial variability. Therefore no assumptions should be made for the zonal gradients: $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$. The meridional gradient of the meridional wind field ($\frac{\partial v}{\partial y}$) is also significant at high latitude as this is where substantial auroral pressure gradients exist. The only remaining derivative is the zonal gradient of the meridional wind ($\frac{\partial v}{\partial x}$).

Burnside et al (1981) made the assumption that, for sufficiently short time intervals, the meridional wind field is stationary in local time so any observed variation is due to the site moving underneath this wind field with the rotation of the Earth. Therefore the meridional wind, $v$, can be sampled at various locations along the zonal direction, $x$. That is:

$$\frac{\partial v}{\partial x} \approx \frac{1}{\epsilon}\frac{\partial v}{\partial t} \tag{3.26}$$

where $\epsilon$ is the tangential velocity of the observing station due to the rotation of the Earth and $\frac{\partial v}{\partial t}$ is calculated from the set of estimates of the uniform component of the meridional wind as a function of time, $\{v_0(t_j)\}$, using the central difference approximation:

$$\frac{\partial v}{\partial t} \approx \frac{v_0(t_{j+1}) - v_0(t_{j-1})}{t_{j+1} - t_{j-1}} \tag{3.27}$$

Burnside et al's (1981) assumption allows information about the wind field's vorticity to be inferred by sampling it from multiple longitudes. This is not possible with single exposures, instead only information regarding the divergence can be inferred. Furthermore, gradients calculated from this method are often very small in comparison with the other gradients (Anderson, 2012). However, this assumption becomes invalid when there is a sudden change in the IMF $B_y$ which affects the thermospheric circulation at high latitudes (Conde and Smith, 1998). This would mean that the wind field seen in local time is not stationary. In fact this argument applies to any time the local wind field is explicitly time-dependent, so would rule out any substorms. Generally, this assumption should be used with caution when used for high-geomagnetic-activity data.

Anderson (2012) makes an alternative assumption, that for a single exposure, the zonal

gradient of the meridional wind field is zero, i.e.,

$$\frac{\partial v}{\partial x} = 0 \tag{3.28}$$

Due to the meridional component showing the least spatial variation, along the zonal direction, this assumption is found, by experience, to give the most plausible wind fields for a single site (Anderson, 2012). By applying Anderson (2012) assumption, $\frac{\partial v}{\partial x} = 0$, the partial derivatives may be fully resolved:

$$\frac{\partial u}{\partial x} = \frac{a_0 + a_2}{R} \tag{3.29}$$

$$\frac{\partial u}{\partial y} = \frac{2b_2}{R} \tag{3.30}$$

$$\frac{\partial v}{\partial y} = \frac{a_0 - a_2}{R} \tag{3.31}$$

The full set of partial derivatives are then fed into the Taylor expansions for $H_x$ and $H_y$ (equations 3.11 and 3.12) and the component of the fitted wind field normal to the line-of-sight is calculated:

$$H^{\perp}(\theta_k, \phi) = H_y \cos\theta_k - H_x \sin\theta_k \tag{3.32}$$

This fitted normal component is combined with the observed LOS component to to give the 'best estimate' of the meridional and zonal components of the actual horizontal wind field:

$$H_x(\theta_k, \phi) = H^{||}(\theta_k, \phi) \cos\theta_k - H^{\perp}(\theta_k, \phi) \sin\theta_k \tag{3.33}$$

$$H_y(\theta_k, \phi) = H^{||}(\theta_k, \phi) \sin\theta_k + H^{\perp}(\theta_k, \phi) \cos\theta_k \tag{3.34}$$

The 'parallel to' and 'normal to' line-of-sight reference frame is implemented in order to retain all the features of the original thermospheric horizontal wind in the LOS direction and include slightly less structure from the fitted field in the perpendicular direction. Even though the fitted component's influence on the best estimate is limited (by taking the perpendicular component), it still improves the accuracy of the field. In effect, it acts to reduce the noise in the final calculation. The method described above can be applied

to each ring of SCANDI separately giving a different set of partial derivatives for each ring. It seems sensible to allow a constant wind gradient over the whole sky so all partial derivatives are averaged. No fitting can be done on the central zenith zone value as there is only one data point, therefore, this is assigned the value of the all-sky mean of horizontal wind vectors.

### 3.11.1 Wind-fit Verification: SCANDI-SuperDARN comparison

To verify the above algorithm, a small study of Svalbard's climatological thermospheric wind field under quiet and active conditions was carried out and compared to our expectations (see chapter one section 1.4.1). In short, we expect the solar pressure gradient force and the ion drag force to be visibly acting in the high latitude thermosphere; the ratios of the two are dependent on the strength and size of the ionospheric convection patterns, however these ratios are not well quantified. The ion drag is thought to be more dominant when the geomagnetic activity is high.

All the clear-sky SCANDI 61 zone line-of-sight wind measurements from 2012-2015 (November-January) are fitted with a horizontal wind field with the algorithm stated in the previous section. The wind fields are binned with MLT and Kp index and averaged. From the geomagnetically active and quiet data two dial plots are made (see figures 3.13 and 3.14). The format of a dial plot has the magnetic north pole at the centre marked by the cross which is labelled with time as 12MLT upwards, 18MLT to the left, 00MLT to the bottom and 06MLT to the right. It is a bird's-eye view of the magnetic north pole of the Earth, with the sun at infinity to the top of the plot. This is a useful layout as this is the same coordinate system in which the polar cap ionospheric plasma convection pattern is usually plotted so makes it easy to compare with ion-drag. Next to each of the dial plots is the statistical sample for both plots showing the number of data points included in each average wind field for each MLT ranging from ∼170-250.

The mean and maximum horizontal wind magnitudes for the quiet data are 86ms$^{-1}$ and 155ms$^{-1}$ whereas for the active data, these are 128ms$^{-1}$ and 247ms$^{-1}$; under quiet times the mean winds are ∼67% of the active wind magnitudes. The active winds are larger, as expected, due to the increased ion-drag accelerating the neutrals. Solar conditions are constant for both data sets so this cannot have an influence on the magnitudes. Both activity dial plots largely show the same general antisunward dawnward flow (top left-to-bottom right 'diagonal' across the plot). There seems to be a very strong component from

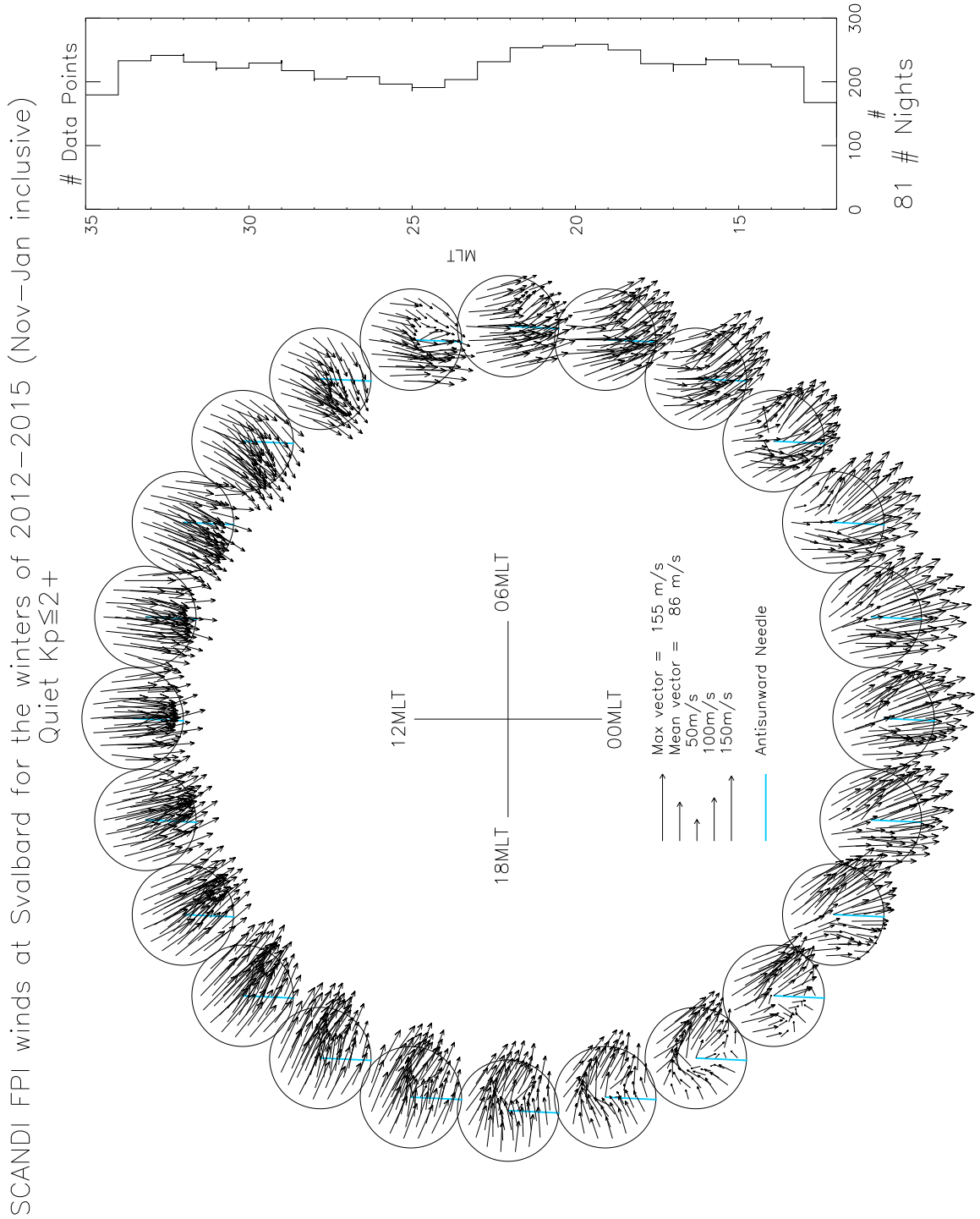


Figure 3.13: Dial plot of SCANDI's climatological wind field with MLT for all geomagnetically quiet data ($K_P \leq 2+$) from 2012-2015 (November to January).

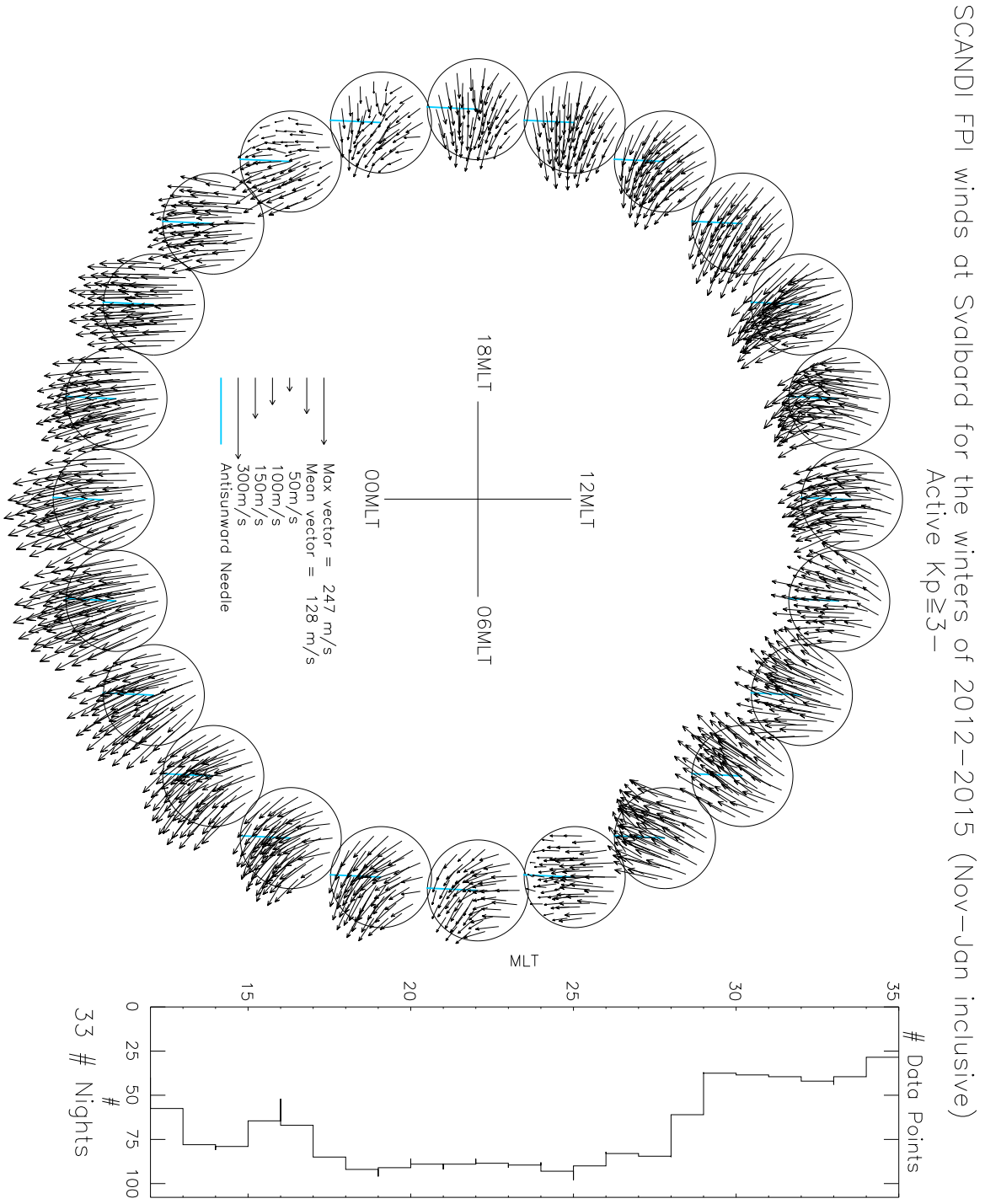


Figure 3.14: Dial plot of SCANDI's climatological wind field with MLT for all geomagnetically active data ($K_P \geq 3-$) from 2012-2015 (November to January).

the pressure gradient force that maximises about 17MLT for Svalbard (14UT) directed straight across the pole. Within the post-midday sector, this component seems to be dominant. The only times it does not seem dominant are in the pre-midday sector, this is likely to be due to influence from the sunward returning plasma drift rotating the vectors inward towards the pole. In the post-midday sector, it seems the sunward returning plasma does not have as much influence because the strong pressure gradient force working in the opposite direction is strongest here and presumably difficult to overpower.

The influence from the ion-drag can be assessed directly using SuperDARN (Chisham et al, 2007) data (courtesy of University of Leicester). Figure 3.15 shows the average SuperDARN plasma velocity vectors plotted in the dial plot coordinate system. The data are binned by Kp just like the SCANDI data. We had no access to 2012-2015 SuperDARN data so the 2003 data are used instead. This is suitable as in 2003 the F10.7 is at the same level as for 2012-2015, i.e., 115<F10.7< 145. This ensures we are looking at the same geomagnetic and solar conditions and the two datasets are comparable. On each plot, the electrostatic potentials are traced out in blue which the ions faithfully follow, mapping out the twin-cell convection pattern. Svalbard's track around the magnetic pole is traced out in red. Where this track intersects the ion vectors are where we must look to assess the direction of ion-drag forcing. The active plasma convection is oriented towards dawnside whereas the quiet plasma convection is only slightly oriented dawnward and is largely aligned from magnetic midday to midnight.

Returning to the issue of the pre-midday deviation from the pressure gradient force direction, we may now see that it is due to the ion drag from the start of the antisunward cross-polar-cap jet, i.e., where the sunward return flow of ions have just turned antisunward and are about to 'shoot across' the pole. The different orientations of the convection patterns show that, Svalbard, under quiet conditions, experiences the start of the antisunward flow at 9-10MLT, whereas, under active conditions, experiences it strongest around 11MLT. A specific example of this on the dial plot is the midday plot. Under quiet conditions, the vectors are pointing straight towards the pole, slightly twisted to the right, mirroring the orientations of the ion flows in the quiet SuperDARN plot. Likewise, under active conditions, the midday plot is mirroring the ions with both vectors oriented $\sim 20^o$ clockwise from the direction straight down towards the pole. The pressure gradient force does not seem to have a direct effect here, although it is likely that some pressure-gradient-induced momentum advection is translated anticlockwise around the dial plot,

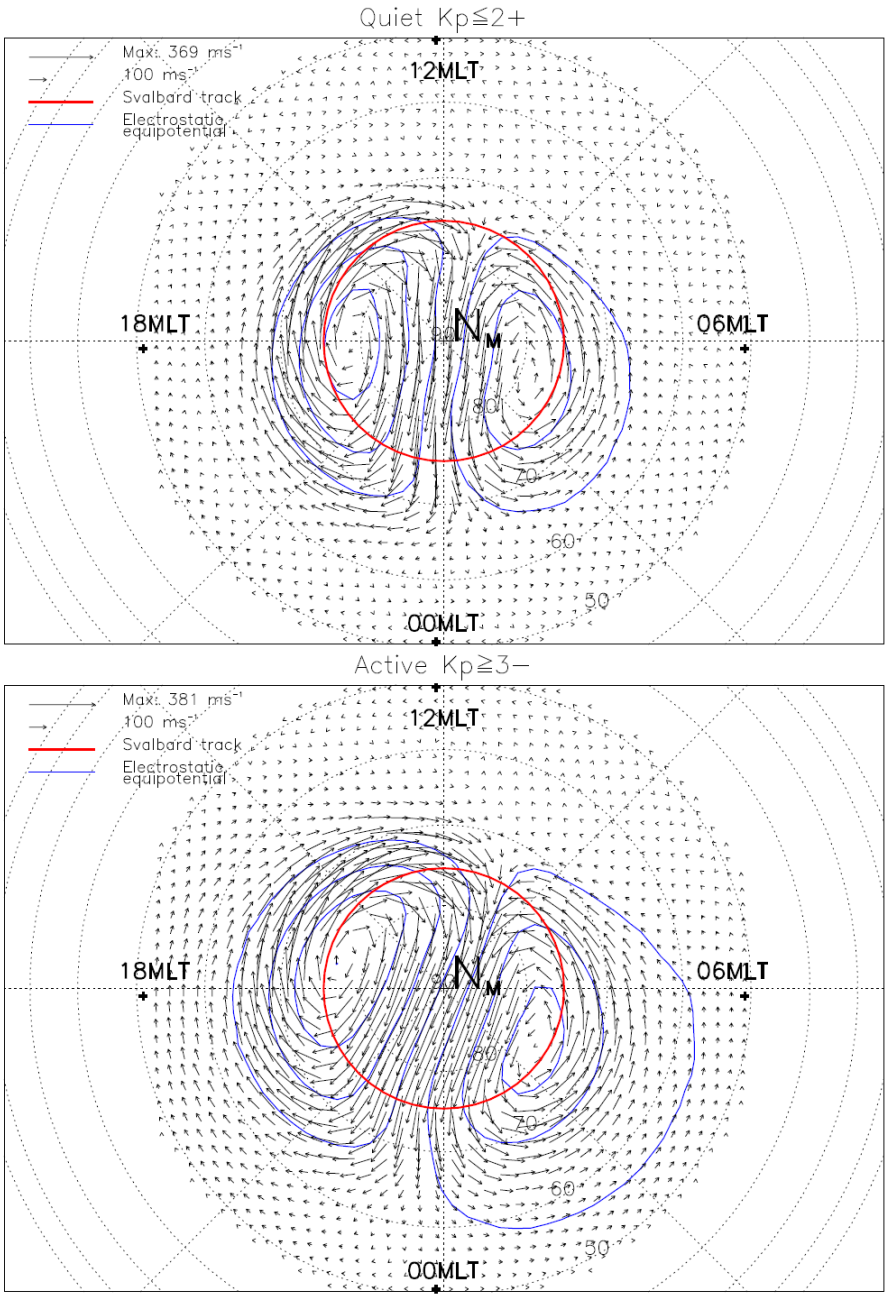


Figure 3.15: Climatology of all SuperDARN 2003 data (November to January) on an MLAT-MLT grid for geomagnetically quiet (top plot) and active (bottom plot) periods. Blue lines show the electrostatic potentials, black vectors show the F region plasma drifts and the red line shows Svalbard's track around the magnetic pole (centre of plot). The data are courtesy of University of Leicester (Chisham et al, 2007).

accelerating the neutral vectors more clockwise.

An obvious disparity between the quiet and active dial plots occurs around magnetic midnight. Under active conditions, the vectors are strongly pointing antisunward, closely mimicking the behaviour of the plasma flow on the SuperDARN plots. Under quiet conditions, however, this antisunward flow is broken up by strong divergences. This behaviour is clearly visible from 23MLT to 03MLT. It is tempting to attribute this to the region of the Harang discontinuity (the general region of which includes a divergent plasma flow at magnetic midnight) lying more poleward under quiet conditions so that it imposes its divergent ion drag forcing on the neutrals. Under active conditions, the Harang Discontinuity region will lie equatorward and Svalbard will thus be expected to lie in the main stream of the cross-polar-cap jet arising from ion drag forcing. However, as seen on the SuperDARN plots, the Harang discontinuity region is located at $\sim 67^o$ latitude under quiet conditions and $\sim 63^o$ under active conditions; Svalbard is always poleward. Although under quiet conditions, on Svalbard's track from 23MLT to 03MLT in the SuperDARN plot, it is possible to see the beginnings of a divergence in the plasma. This is not evident within the SuperDARN active time plot. In the active time plot, as predicted, Svalbard lies too far in the bulk of the antisunward plasma jet for any divergence to be experienced.

Another point to note for the quiet time dial plot compared to the active one is the larger convergence in the pre-midnight sector (between 19-22MLT specifically) and the larger divergence in the dusk sector (between 04-07MLT specifically). In other words, the flanks of the thermospheric convection seem to be more disordered under quiet conditions, whereas the active conditions are characterised by far 'smoother' flows. If we refer to the SuperDARN plots we may see that under quiet conditions Svalbard's track is more towards the edges of the convection cells, where the average sunward return flow of ions reach magnitudes of $150 \mathrm{ms}^{-1}$. Under active conditions where Svalbard's track goes through the centre of each convection cell, the magnitudes drop and appear approximately 50% smaller. This reduced ion speed appears to have little or no drag effect on the neutrals, hence the smooth active wind field (appearing largely controlled by momentum advection from the pressure gradient force). Under quiet conditions, in the premidnight sector the sunward flowing ions seem to be battling with the pressure gradient force (or the resulting momentum advection); this explains the convergence. In a similar vein, the quiet-time dawn sector divergent winds seem to be due to the battling of the sunward return-flowing ions and the momentum advection from the pressure-gradient forcing. Ion drag is not more

powerful under quiet conditions, it is just that Svalbard is sampling the more disruptive part of it due to the shrinking of the auroral oval. The active wind field still experiences great forcing from the ion drag, however this is exhibited to Svalbard mostly as the cross-polar-cap jet and this tends to blend in smoothly with the pressure gradient forcing.

In conclusion, all aspects of both the active and quiet dial plot can be explained with known thermospheric wind behaviour and the help of the SuperDARN plasma convection plots. As the wind fields are complying with our expectations, we can be confident that our treatment of the SCANDI raw data and subsequent wind-fitting is yielding results of reasonable accuracy and leads to meaningful conclusions.

# Chapter 4

# Vertical Winds

Little work has been done on vertical winds because they are notoriously difficult to measure. With typical speeds of 10ms$^{-1}$ or less, they are an order of magnitude smaller than their horizontal counterpart and until recently the FPI detectors lacked the sensitivity to measure them with enough signal so they could be distinguished from the noise. With the invention of EMCCD chips, the sensitivity of the FPIs increased dramatically and this allowed integration times as low as ten seconds rather than the few minutes that were required with the previous photometer detectors. Vertical winds can now be measured at an unprecedented precision and accuracy allowing short-lived larger magnitude winds to be detected where previously these would be smoothed out by the longer integration times.

On a global scale the vertical wind is driven by solar-heated pressure gradients which cause the atmosphere to expand and contract diurnally and thus produce tides which never contribute more than a few ms$^{-1}$ wind magnitude (Smith, 1998). However, at high latitudes geomagnetic heat sources become important, such as Joule and particle heating, which can also set up diurnal variations as the site makes its daily transit underneath the geomagnetic heat source. These diurnal movements of the atmosphere are assumed to never exceed ∼50ms$^{-1}$ (Aruliah et al, 1994; Rees et al, 1984a; Smith and Hernandez, 1994) although on many occasions extremely large vertical wind events greater than 200ms$^{-1}$ have been reported at auroral latitudes, often associated with large thermospheric temperature enhancements with increases of 100-500K (Spencer et al, 1982; Crickmore et al,

1991; Innis et al, 1996; Kurihara et al, 2009).

Auroral energy deposition enhances the neutral temperature directly causing a sudden rise in local temperature which causes these vertical winds to be generated. Changes in vertical wind seem to respond instantaneously to changes in the heating rate (Kurihara, 2009). Oyama et al (2008) derive neutral vertical wind using EISCAT data and find magnitudes exceeding 30ms$^{-1}$ accompanied by an auroral heating event. Guo and McEwen (2003) observe upward vertical wind during and following a sustained aurora above their polar cap site, suggesting that heating due to particle precipitation is sufficient to sustain upwards flow. Rees et al (1984a) identify a source of intense local heating with a period of 15-30mins in the magnetic midnight region driving rapid upward vertical motion (50-100ms$^{-1}$) accompanied by intense magnetic disturbances, evidence of the MIT system at work.

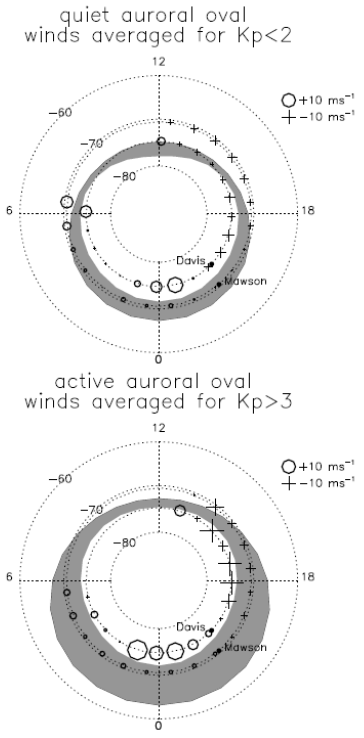


Figure 4.1: South pole distribution of positive and negative vertical winds with respect to the auroral oval at geomagnetically quiet (top) and active (bottom) conditions (Greet, 2002)

The primary focus of this chapter is the apparent relationship between relative location with respect to the auroral oval and vertical wind direction and magnitude. Generally, it is reported, on the poleward edge of the auroral oval there are more often upwards vertical

winds whereas the downward winds are more commonly observed at the equatorward boundary (Spencer et al, 1982; Crickmore et al, 1991; Innis et al, 1996). This phenomenon is investigated more explicitly by Greet et al (2002) who developed MLAT-MLT maps (figure 4.1) showing the average vertical wind vectors distributed around the auroral oval. In figure 4.1 the size of the symbol represents the speed of the vertical wind which reveals that the larger magnitude winds are found on the poleward side of the oval compared to those within the oval. This chapter will lead onto chapter 5 as the Svalbard midnight vertical winds required a dedicated investigation into their source.

## 4.1 Auroral Oval Position

### 4.1.1 Models

Before commencing the investigation, we review two auroral oval models: Feldstein and Starkov's (1967) and Hardy et al's (1987) models. The auroral oval is the primary geomagnetic heat source and there have been many reports of large vertical winds in relation to it (Kurihara et al, 2009; Oyama et al, 2008; Greet et al, 2002; Guo and McEwen, 2003; Rees et al, 1984a), so it is useful to understand its spatial relationship with the two observing sites. The position and size of the auroral oval are highly variable due to their dependence on the geomagnetic activity which is also highly variable throughout the night. Therefore we must be cautious when comparing a local measurement of vertical wind to statistically averaged auroral oval models (Feldstein and Starkov, 1987; Hardy et al, 1987). Furthermore, in the following review we should bear in mind that any discussion of the degree to which the site makes an excursion into the auroral oval translates to the probability that the auroral oval is found overhead at the site.

Feldstein and Starkov (1967) determine the average position and shape of the auroral oval belt by combining IGY-IGC all-sky camera observations from many stations and binning them into three different levels of activity displayed in figure 4.2. The two bold circular lines on each map represent our sites with the inner red one representing Svalbard and the outer blue one representing Kiruna. According to their model, under quiet conditions Kiruna ($\sim 65^o$ magnetic latitude) is always equatorward of the contracted auroral oval. For more disturbed geomagnetic conditions, Kiruna's track dips into the equatorward edge of the auroral oval on the nightside. The time spent and 'depth' at which Kiruna is inside the auroral oval increases with activity.

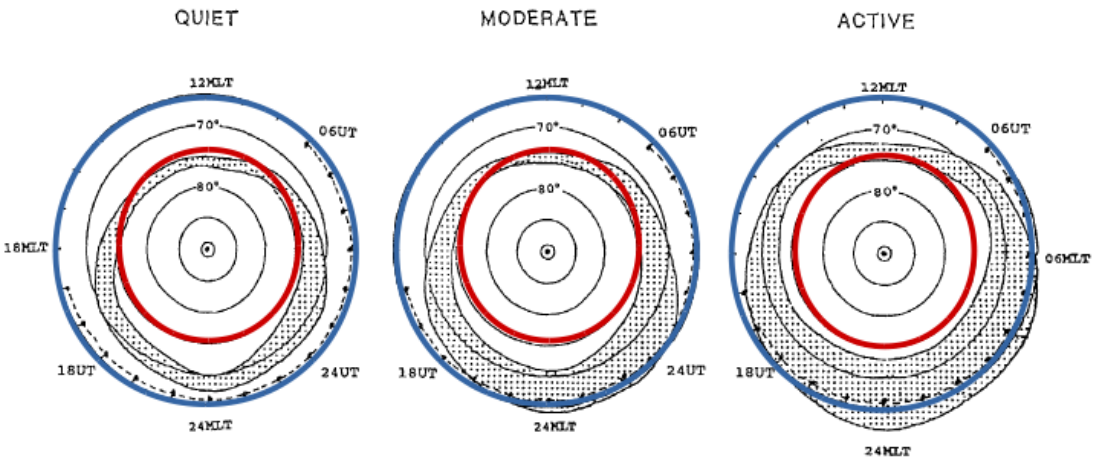


Figure 4.2: Feldstein and Starkov (1967) show three MLAT-MLT maps for quiet, moderate and active levels of geomagnetic activity. The bold circles have been added to the plots to show Svalbard's (red) and Kiruna's (blue) orbits around the magnetic pole for their respective magnetic latitudes of 75.43°N and 64.89°N.

Svalbard ($\sim 75^o$ magnetic latitude), however, has the opposite relationship to the auroral oval: it is always poleward of the auroral oval at all times except for when it passes underneath on the day-side. Under more disturbed geomagnetic conditions, the auroral oval moves further away from Svalbard (equatorward) on the dayside however the midnight bulge moves towards Svalbard on the night-side. This results in Svalbard spending less time underneath the auroral oval around midday and being closer to the auroral oval on the nightside, but not directly underneath.

Hardy et al (1987) carried out a statistical study with the DMSP satellite data to produce global statistical maps of average Pedersen and Hall conductivities as a function of geomagnetic activity ($K_P$ index). The maps from $K_P$=0-4 is shown in figure 4.3. The conductivities are good representations of the auroral oval as auroral arcs are supported by current systems, composed of both parallel and perpendicular components. Parallel current corresponds to particle precipitation which may enhance local conductivity through ionisation. In figure 4.3 the reported relationship between Kiruna and the auroral oval is consistent with Feldstein and Starkov (1967)'s model: equatorward except under moderate and active conditions when its track passes underneath on the nightside; the more active the deeper are the excursions into the auroral oval at Kiruna.

Svalbard's relation to the auroral oval in Hardy et al's (1987) model is similar to Feldstein and Starkov (1967)'s with its track being poleward of the auroral oval under quiet conditions and the midnight bulge moving to a closer proximity under active conditions.

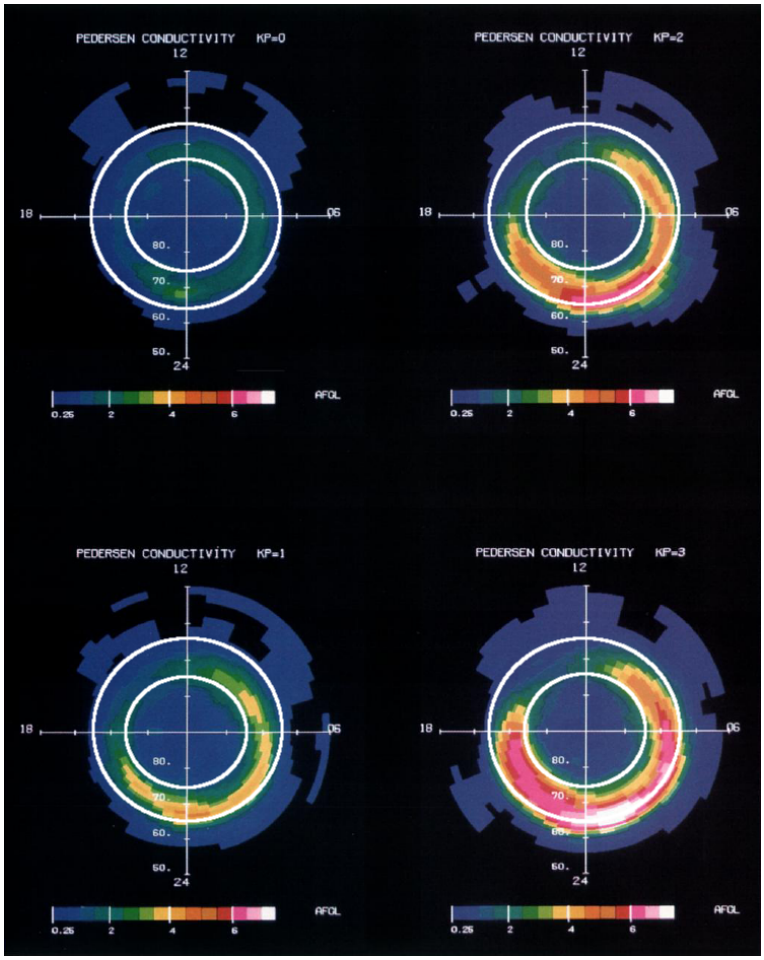


Figure 4.3: Hardy et al (1987) Pedersen conductivity MLAT-MLT maps for four different activity levels. The bold white circles have been added to the plots to show Svalbard's (inner) and Kiruna's (outer) orbits around the magnetic pole.

In Feldstein and Starkov's (1967) model, the only time Svalbard passes under the auroral oval is on the dayside however in Hardy et al's (1987) model around noon there is a gap in the auroral oval. This could be due to the relatively low energy of the soft particle precipitation around noon creating a visible aurora but not increasing the conductivity enough to show up on the map, relative to the effect of the hard night-time precipitation.

Hardy et al's (1987) entire auroral oval, in particular at the nightside, is latitudinally 'fatter' under more disturbed conditions. The nightside oval expands further equatorward to $<60^o$ latitude and extends further poleward to $>75^o$ latitude, approaching Svalbard's track. Feldstein and Starkov (1967)'s auroral oval, under more disturbed conditions, barely reaches $60^o$ at its equatorward edge or $75^o$ at its poleward edge; it is much latitudinally thinner. This disparity could be due to the auroral oval being defined by conductivity in Hardy et al's (1987) model and by the more conventional visible light in Feldstein and

Starkov's model (1967). The visible light is produced by particle precipitation exciting atmospheric species and causing emissions. These particles are also responsible for ionising the atmosphere so that there is also relatively high conductivity in the auroral region. However the conductivity enhancement spreads further latitudinally than the visible light, which is likely to be due to the electric field configuration around the auroral arc to satisfy current continuity (Kosch et al, 2011). Another reason may be that the remote ASCs used to measure the visible light do not detect the fainter emissions around the edge of the arcs, whereas the in situ DMSP satellite has the greater sensitivity to detect all incoming particles.

In conclusion, these models have shown that Svalbard, under all geomagnetic conditions, can be assumed to be poleward of the auroral oval all the time, apart from at midday, when it is typically underneath. However, at more geomagnetically active times Svalbard spends less time underneath the auroral oval around midday. Kiruna, at geomagnetically quiet times, is always equatorward of the auroral oval, coming in very close proximity to it around midnight. Under more active conditions, Kiruna is underneath the auroral oval at midnight; the degree to which Kiruna is engulfed is dependent on the level of activity. In the next subsection we can test these models by comparing with the FPI data.

### 4.1.2 FPI Intensities

Confirming the FPIs spatial relationships to the auroral oval is achieved by considering their average emission intensity throughout the day and comparing the trend with Kiruna's expected behaviour, which, at magnetic latitude $65^o$, is known to be an auroral oval site at midnight. The magnitudes cannot be compared between the sites as the different FPI detectors are not cross-calibrated and it would be futile to do so. There are too many parts within an FPI instrument which affect the detected intensity over time (see section 2.7 in chapter 2), hence why the intensity measurements are relative.

Figure 4.4 shows the average Svalbard (top two) and Kiruna (bottom two) red line relative emission intensities as a function of UT. The intensity data is binned with UT and an arithmentic mean is taken of each bin. The bottom of each of these two sets of plots are the numbers of data points included in each average. The green lines represent quiet data where $0<K_P \leq 2+$ and the purple lines represent active data where $K_P > 2+$ (typical $K_P$ boundaries chosen, e.g, Aruliah et al, 1995). The 2007 year has been excluded from the Svalbard dataset because the relative intensity of that year is so high (see figure 2.7),

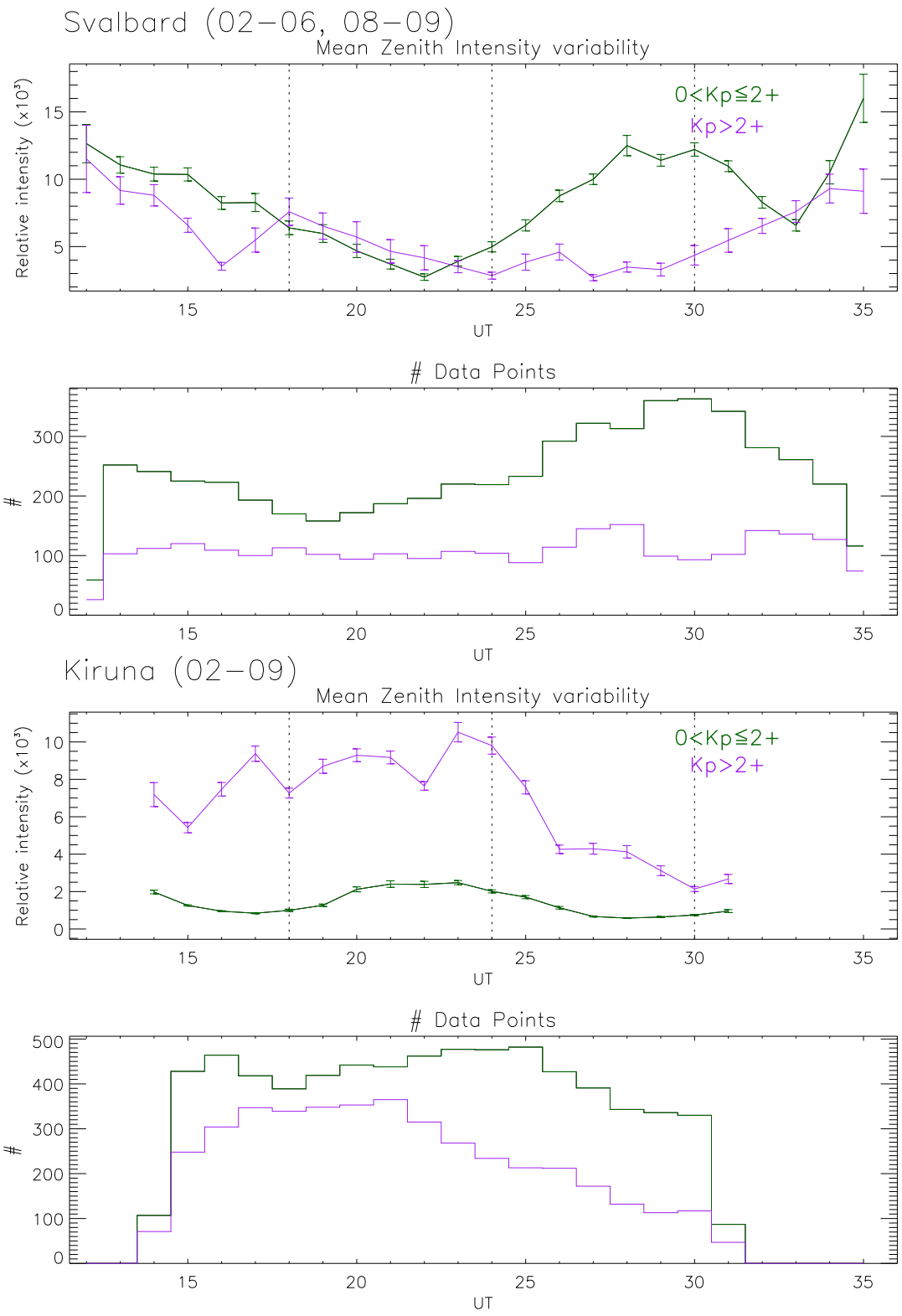


Figure 4.4: Statistical averages of Svalbard's and Kiruna's relative intensity emissions varying with UT. The top two plots show Svalbard data and the bottom Kiruna. For each site the top plot shows the average value variation and the plot underneath shows the number of data points used in each average. The data are also split by geomagnetic activity represented by the colour of the lines: green represents quiet data where $0<K_P \leq 2+$ and purple represents active data where $K_P > 2+$.

it would too heavily bias the average. The error bars display the standard errors of the averaged samples. These are consistently small compared to the trend variations. This shows the calculated averages are accurate estimates of their population means and show a good statistical representation of the emission behaviour.

Firstly considering the Svalbard quiet variation, it has a minimum centred roughly about magnetic midnight, indicative of being outside the auroral oval, and a maximum about magnetic dawn (28UT). This maximum is not evident in the active dataset. In fact, compared to the quiet trend, the active data show a reduced intensity throughout the whole night apart from a marginal increase for a few hours around 20UT. This suggests that generally the auroral oval moves further away (equatorward) from Svalbard under active conditions. However, at midnight, the auroral oval may grow both poleward and equatorward. Both active and quiet datasets show an increase in intensity at the edges of the day towards 12UT, indicative of the soft cusp precipitation.

Secondly, considering the Kiruna quiet trend, there is a clear maximum near magnetic midnight attributable to the intensity enhancement of the midnight auroral region. It is a smooth trend indicative of quiet geomagnetic activity. The active trend is very different from the quiet one with intensity values 3-5 times higher. The trend is also less smooth, indicative of the variability of the active auroral oval.

To summarise, around magnetic midnight, Kiruna experiences a maximum in emission intensity (larger under active conditions), whereas, Svalbard experiences a minimum (smaller under quiet conditions). Conversely, at magnetic midday Kiruna experiences its lowest emission intensities, whereas Svalbard experiences its largest. These findings confirm the relationships inferred from the two models reviewed in the previous subsection. Since the energy deposition within the auroral oval has been known to drive upward vertical winds (Guo and McEwen, 2001; Kurihara et al, 2009; Oyama et al, 2008), we can use these auroral oval proximity relationships when analysing the vertical winds in the following sections and chapter 5.

## 4.2 Diurnal Vertical Wind Variation

Diurnal vertical wind trends introduced by high-latitude geomagnetic sources are found to overshadow the gentle diurnal trend produced by solar heating. Two of the early contributors to the study of geomagnetically sourced diurnal vertical wind trends are

Aruliah et al (1994) and Rees et al (1984a) who used the Kiruna FPI and Svalbard FPI respectively. These are the same two instruments that produced the data for the following study. However, many (possibly all) of the FPI components have been changed over the years.

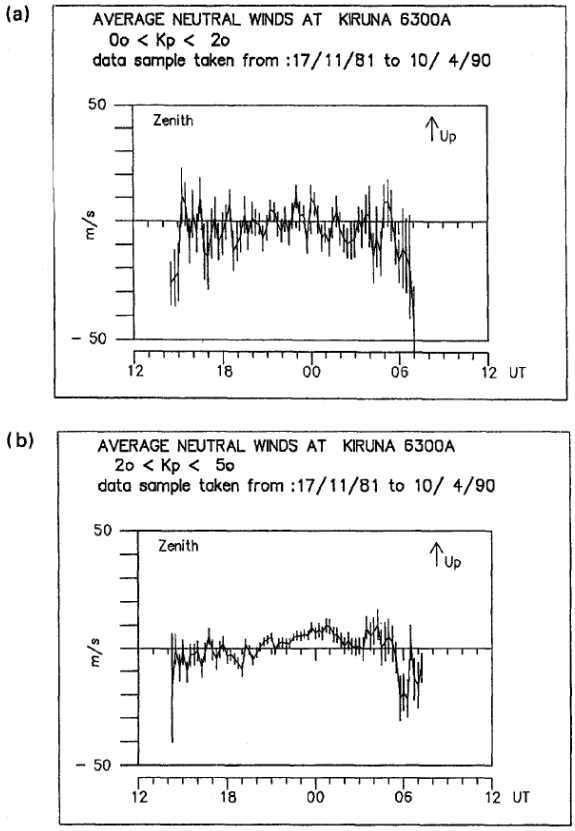


Figure 4.5: The mean vertical wind in the thermosphere above Kiruna as a function of UT split by geomagnetic activity during the period November 1981-April 1990 (Aruliah et al, 1995). Sub-plot a) is for $0<K_P<2o$ and sub-plot b) is for $2o<K_P<5o$.

Aruliah et al (1994) carry out a statistical analysis using eight seasons of the Kiruna FPI data to show a dependence of thermospheric vertical wind on geomagnetic activity, solar cycle, UT and season. Figure 4.5 shows the mean vertical wind in the thermosphere above Kiruna as a function of UT split by geomagnetic activity during the period November 1981-April 1990 (Aruliah et al, 1995). Sub-plot a) is for $0<K_P<2o$ and sub-plot b) is for $2o<K_P<5o$. It is found that active nights show a more sustained period of vertical wind, sometimes up to 6 hours, with an average amplitude of $10ms^{-1}$. For the quiet nights the vertical wind oscillations are shorter in period and have smaller amplitudes, even though they can sometimes reach $10ms^{-1}$.

In figure 4.6 Rees et al (1984a) show that there is a diurnal recurrence of a vertical

wind pattern seen in the F region polar cap winds at Svalbard over seven days in 1980; a downward wind on the dayside and an upward one on the nightside. A similar diurnal pattern is reported by Smith and Hernandez (1994) using upper thermospheric data from the South Pole situated at a similar geomagnetic location. Rees et al (1984a) explain this downward flow by saying it was caused mainly by a divergence in the horizontal motion, however the nightside upward wind is explained by heating and dynamically driven horizontal flow.

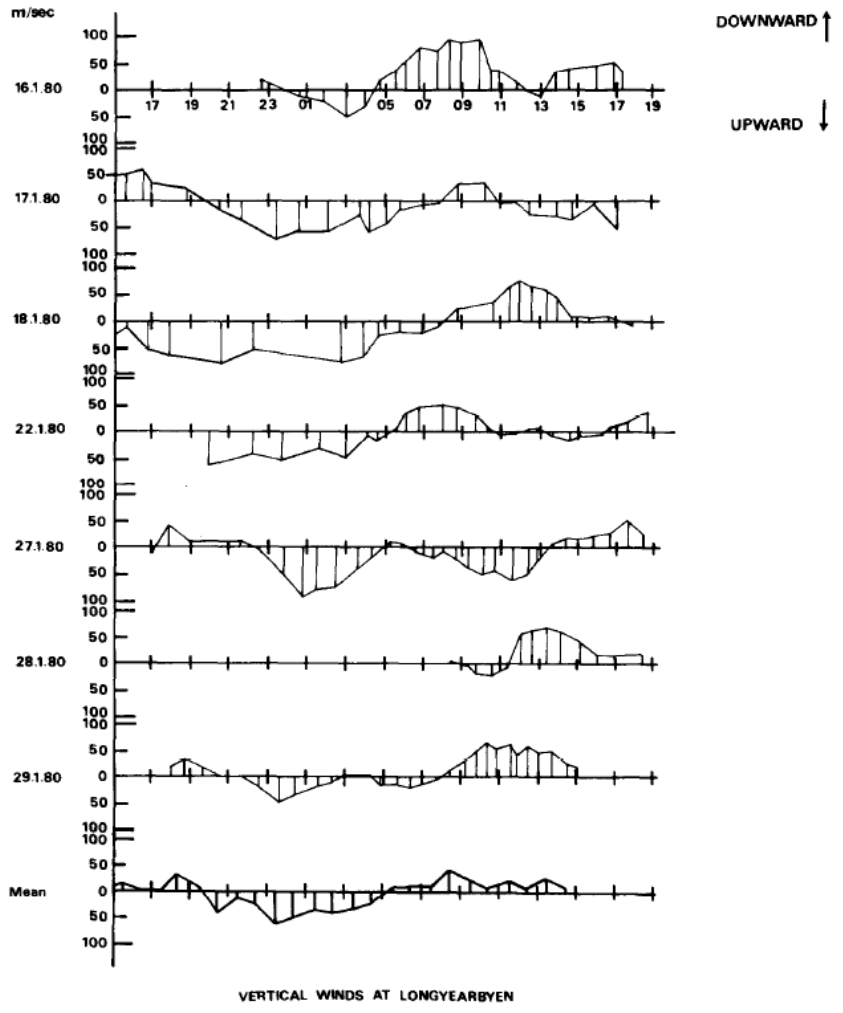


Figure 4.6: Vertical wind observations from 7 nights in January 1980 from Svalbard FPI 630nm data (Rees et al, 1984a). The figure shows the thermospheric wind is systematically upward for a period of several hours at night-time (maximum magnitude reaching 100m/s) followed by systematic downward wind at daytime.

The combined Svalbard and Kiruna database used in this study is accompanied by ASC data from 2002-2009 and is large enough to show significant differences in the behaviour of the thermospheric wind in the polar cap and at auroral latitudes under different

geomagnetic conditions. The introduction of the ASC to the sites (1999 for Svalbard and 2002 for Kiruna) means there is unprecedented confidence in the absence of cloudy data within our statistical sample.

Figure 4.7 compares average vertical wind data from periods of low activity ($K_p$ <2o) with periods of moderate to high activity (2o<$K_P$ <5o) for both Svalbard (sub-plots a)) and KEOPS (sub-plots b)) observing sites. Positive values correspond to an upward wind. The mean vertical winds are plotted every hour (MLT) and the error bars show the standard deviation. Underneath the average vertical wind plot for each site are plots displaying the number of data points used in each hourly average. These clearly show fewer data points at the start and end of the 24-hour period compared with the intermediate times, which is an artefact of the seasonal variability of night-time observing hours. The statistical sample for each plot has adequate data points for the calculated standard error to be never more than 1ms$^{-1}$. The standard deviation, however, ranges from 5-40ms$^{-1}$ for Svalbard and 1-12ms$^{-1}$ for KEOPS, showing Svalbard's winds to be more variable than those at Kiruna.

Firstly, considering Svalbard's vertical wind variation (sub-plot a)i)), there is a systematic diurnal pattern showing strong upward winds at night-time and downward winds at daytime under both active and quiet geomagnetic conditions; consistent with Rees et al's (1984a) trend. Both trends peak around magnetic midnight but with a larger amplitude at quiet-time, nearly 40ms$^{-1}$, compared with active-time, just above 20ms$^{-1}$. Rees et al's (1984a) peak wind magnitude reaches 50ms$^{-1}$ which is ∼10-30ms$^{-1}$ larger than the data presented; this is probably due to the small statistical sample used by Rees et al (1984). There is no upward wind signature of the cusp near noon even though, statistically, it is expected to pass right overhead Svalbard's zenith. This is probably due to the cusp mapping to too spatially localised a region of the atmosphere for the FPI to frequently detect.

Secondly, considering Kiruna's vertical wind variation (figure 4.7 b(i)), during periods of both quiet and active geomagnetic activity the mean vertical wind appears to show a systematic variation throughout the night, however there is a marked difference between the two trends. For quiet times there is a small downward wind ∼5 hours wide around magnetic midnight reaching -10ms$^{-1}$ and at daytime there is a small upward wind reaching 10ms$^{-1}$. This trend is inconsistent with Aruliah et al (1994) who report quiet time winds varying irregularly about zero throughout the night. We must bear in mind that Aruliah

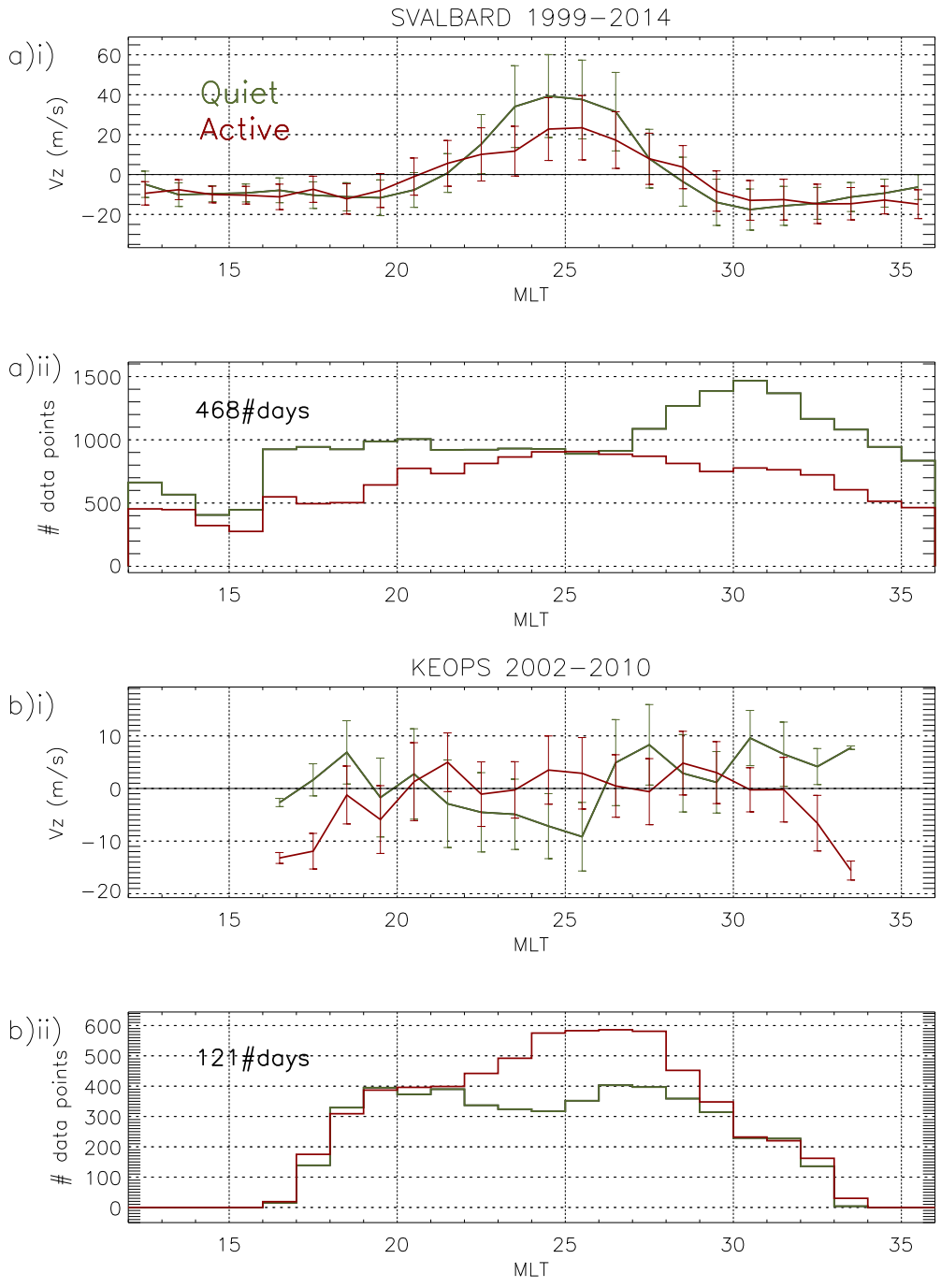


Figure 4.7: Average vertical wind variation with MLT for a)i) Svalbard and b)i) Kiruna for all data accompanied by ASC data with their hourly statistical sample number plotted underneath each in the ii) plots. The green line corresponds to quiet data ($K_P$ <2o) and the red line corresponds to active (2o<$K_P$ <5o)

et al (1994) inferred clear-sky data from the wind signals as there were no ASC images; this could have introduced errors to the wind velocities. The discrepancy is most likely to be due to the quiet-time signals being relatively weak for the far less sensitive detectors of the 1980s, thus contributing error.

For Kiruna's active times, the vertical winds show a weak oscillating upward wind (not significant oscillations as its oscillating within the error bars) around midnight reaching 4-5ms$^{-1}$ and then downward winds towards the daytime reaching -15ms$^{-1}$. The general trend is consistent with that reported by Aruliah et al (1994); similar magnitudes and the same reasoning applies, i.e., the period of upwelling is matched by the period in which the auroral oval lies overhead.

The Kiruna winds' magnitudes are relatively small compared to the Svalbard winds'. Although consistent with previous work (Rees et al, 1984a; Aruliah et al, 1994), this seems counter-intuitive in theory. The likely generation mechanism for high latitude winds is auroral in origin. It has been shown in the previous section that Kiruna is underneath the auroral oval more often than Svalbard so the winds there would be expected to be larger and more persistent than Svalbard's. This is not observed to be the case.

Another odd finding in this study is that Svalbard's maximum midnight vertical wind speed is larger for quiet times when the site is found to be further from the auroral oval (see section 4.1). Along the same vein, it seems odd that Svalbard's largest vertical wind values are centred about midnight when this is the time the site is expected to be furthest from the auroral oval.

In summary, Kiruna's vertical winds are mostly consistent with Aruliah et al's (1994) findings. The vertical wind trend seems plausible considering the sites location with respect to the auroral oval, and, their small magnitudes indicate compliance with hydrostatic equilibrium within the auroral oval. Svalbard's vertical winds are consistent with Rees et al (1984a); however, their relationship with the auroral oval seems counter-intuitive, and, their large magnitudes do not seem reasonable (considering the polar cap is a relatively geomagnetically quiet region compared to the auroral oval) indicating violation of hydrostatic equilibrium in the polar cap. These issues are investigated further in the next sections.

## 4.3 Vertical Wind Frequency Distributions

It is insightful to consider all the vertical wind data at once in a frequency distribution. The only loss of information comes from the binning of the data into vertical wind bins but, provided that the bins are narrow enough, this effect is negligible. It is also important that the bins are not so narrow that the data-point count is too low and a sensible distribution shape cannot emerge. The Freedman-Diaconis rule is employed to select the optimum bin size: 2ms$^{-1}$ for Svalbard and 3ms$^{-1}$ for Kiruna. The equation for this rule is:

$$Bin\_size = 2 \times IQR(x) \times n^{-1/3} \tag{4.1}$$

where IQR(x) is the interquartile range of the data and n is the number of observations in the sample x. This rule aims to minimise the sum of squared error between the relative frequency histogram and an arbitrary probability density (Keen, 2010).

Figure 4.8 shows the resulting frequency distribution of all the raw vertical wind data from Svalbard (navy) and Kiruna (pink) FPIs that have corresponding all-sky camera-derived cloud records. The plot spans -500ms$^{-1}$ to 450ms$^{-1}$ to display all outliers.

For Svalbard the vertical wind frequency peaks at -11ms$^{-1}$, and the distribution is upward skewed (skewness value, $\gamma$=1.5). For Kiruna the vertical wind frequency peaks at 5ms$^{-1}$ and the distribution is roughly symmetric ($\gamma$=-0.03). Its distribution counts are relatively small due to there being fewer night-time observing hours at a lower latitude and fewer all-sky-camera-derived cloud records to pair with the data, i.e., 8 years instead of Svalbard's 14 year run. It is not reasonable to cross-compare these distributions as they are not normalised and they each include different ranges of observing years. A comparison between Svalbard's and Kiruna's vertical wind distributions is carried out in the following section (4.3.1).

The purpose of showing these distributions is to highlight the excess of extreme positive vertical wind anomalies for Svalbard. Any extreme wind magnitudes >5-sigma away from the mean of the distributions are considered anomalies and are plotted on the distributions as crosses (navy for Svalbard and pink for Kiruna). Coincidentally, both sites' 5-sigma limits are calculated to be approximately ±175ms$^{-1}$. The Kiruna distribution displays only one anomaly at -395ms$^{-1}$, whereas Svalbard has many examples spanning -530ms$^{-1}$ to 415ms$^{-1}$, with most of them being positive. There are also many vertical winds <175ms$^{-1}$ in magnitude which could still be anomalous seeing as high latitude vertical winds are

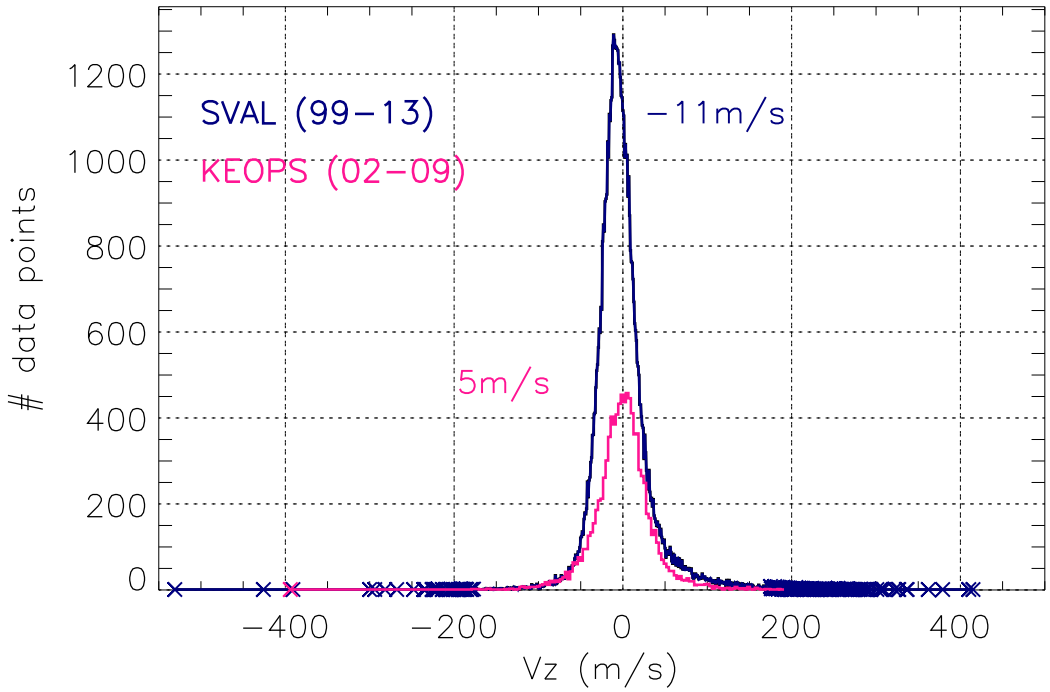


Figure 4.8: Frequency distribution of all raw vertical wind data that is accompanied by ASC data, i.e., for Svalbard 1999-2015 (navy) and for KEOPS 2002-2009 (pink). The Svalbard distribution peaks in the -11ms$^{-1}$ bin and KEOPS' peaks in the +5ms$^{-1}$ bin. Outliers are marked as crosses for any vertical wind bin which lies outside the 5-sigma boundaries, i.e., $|Vz|$ >175ms$^{-1}$ for both sites (which both share a standard devation of 35ms$^{-1}$).

generally ±50ms$^{-1}$ in magnitude.

The excess of extreme upward vertical winds evident in figure 4.8 is surprising, particularly so for the Svalbard polar cap site; it would take a vast amount of energy to drive these magnitudes. Innis et al (1999) estimate a power of $6\times10^{9}$W to drive upwellings of this size over a typical upwelling area of $6\times10^{11}$m$^2$ (where the global Joule and particle heating powers are estimated to be 7-10$^{10}$W (Roble et al, 1987)). These anomalous data points are likely to be the reason for the relatively large, average-magnitude midnight upwelling evident in the Svalbard vertical wind variation in figure 4.7.

The asymmetry and excess of positive extreme vertical winds make these distributions compelling and so this line of investigation is developed in the coming chapter. The distributions in the following section (figure 4.9) allow a cross-comparison between the two sites.

### 4.3.1 General Svalbard-Kiruna Comparison

Figure 4.9 displays Svalbard (navy) and KEOPS (pink) histograms which may be cross-compared. Data are from the same years (2002-2009) and from the months of November, December and January when Svalbard is under 24 hour darkness so that the FPI can observe continuously without interruption. Kiruna also experiences its maximum daily duration of darkness at approximately 18 hours. Thus, the statistical sample is maximised whilst the bias towards any particular hour is minimised.

Sub-plot a) displays the frequency distributions spanning bins -150ms$^{-1}$ to 150ms$^{-1}$ excluding all anomalies out of range which are likely to be spurious (>4-sigma away from the mean values). Kiruna's distribution's peak height is scaled to be in line with Svalbard's with its axis in pink on the right hand side. Svalbard's is on the left in black. The peak vertical wind bins are labelled for each which are identical to the ones in figure 4.8. Sub-plots b) and c) show Svalbard and Kiruna's distributions separately with an attempt to fit a Gaussian peak with the mean ($\mu_{FIT}$) and standard deviation ($\sigma$) stated. Under this the kurtosis ($\kappa$) and skewness ($\gamma$) of the distribution are stated. The kurtosis quantifies how much the data are distributed away from its mean or how 'pointed' the distribution is. Any distribution with kurtosis >3 (<3) is considered leptokurtic (platykurtic), i.e., more (less) pointed than a Gaussian. A large kurtosis indicates that extreme events have a sizeable probability of occurrence.

The best-fitting Gaussians over-plotted in sub-plots b) and c) highlight the departure from Gaussianity of both distributions. The Kiruna distribution fits the Gaussian better with a goodness of fit $\chi^2$ value of 18 compared to Svalbard's value of 55 (reduced $\chi^2$ statistic is used, $\chi^2 = \frac{1}{DOF}\Sigma\frac{(x_O - x_M)^2}{\sigma^2}$ where $x_O$ is the observed data, $x_M$ is the theoretical data and DOF represents the degrees of freedom, equal to N-n-1 where N is the number of observations and n is the number of fitted parameters). However, both fits are poor which is mainly due to the wide tails of the distributions; both distributions are leptokurtic ($\kappa > 3$). The Svalbard distribution b) is particularly leptokurtic, with $\kappa$ =11.7, due to its extremely wide positive tail. Its negative tail will not be contributing to the high kurtosis as it fits the Gaussian. This asymmetry is reflected in the skewness value of 2.49 representing the heavy positive bias. The skewness of the Kiruna distribution c), value of -0.01, shows a very slight negative skew; almost perfect symmetry. There is a minimal level of kurtosis on either side of the distribution, with a $\kappa$ =5.3.

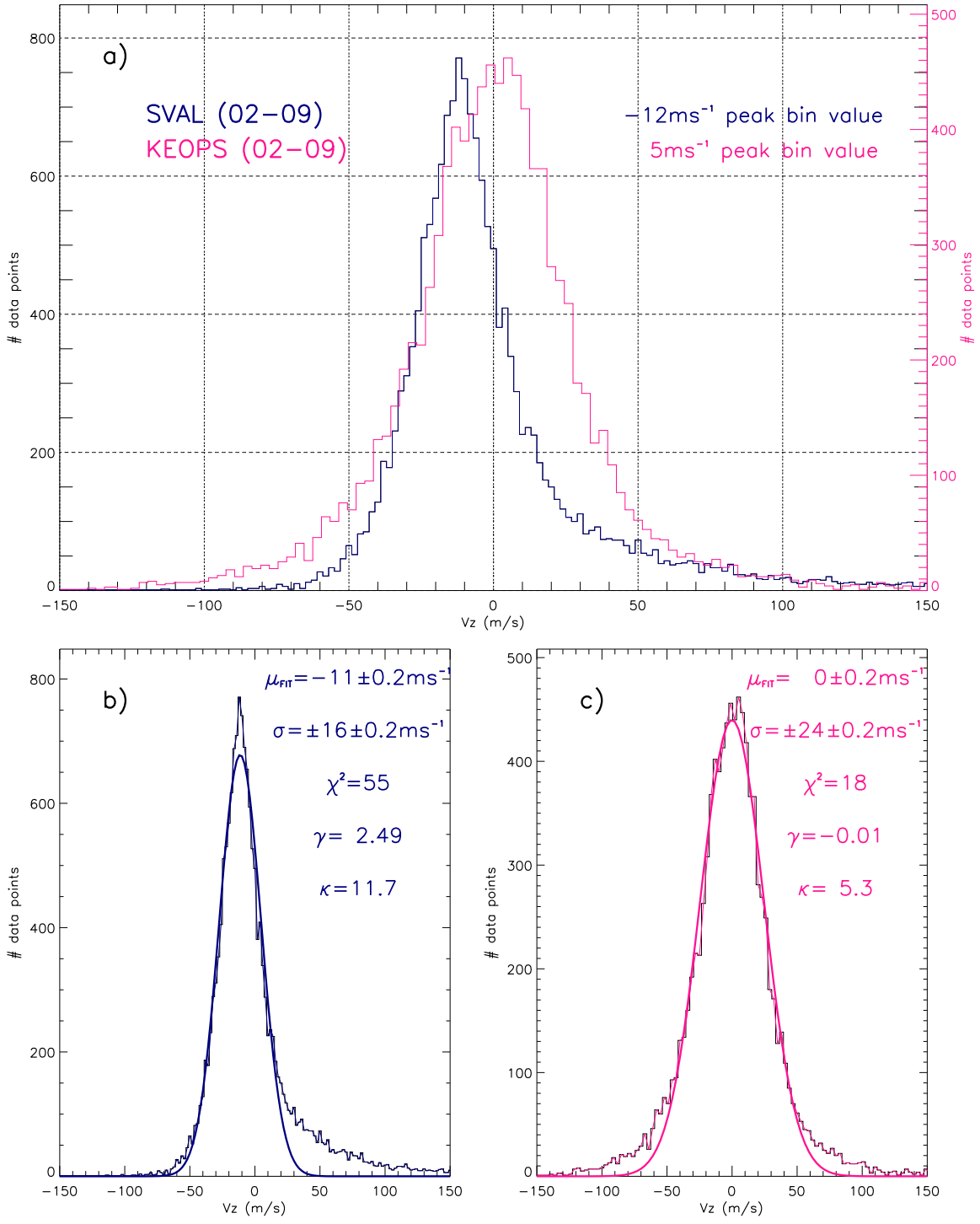


Figure 4.9: Frequency distribution of raw vertical wind data for KEOPS (pink) and Svalbard (navy) from 2002-2009 November-January (all accompanied by ASC data). In sub-plot a) Svalbard data's axis is in black on the left and KEOPS' is in pink on the right scaled so that the distributions are the same height. Sub-plots b) and c) are the Svalbard's and KEOPS' distributions, respectively, plotted separately. An attempt to fit a Gaussian is over-plotted on each and labelled to the side are: the mean of the fit $\mu$, the standard deviation of the fit $\sigma$, the kurtosis of the distribution $\kappa$ and the skewness of the distribution $\gamma$.

The Kiruna distribution's fitted Gaussian mean value is 0ms$^{-1}$ with an error of $\pm$0.2ms$^{-1}$ (small compared to standard deviation). This Gaussian mean values is 5ms$^{-1}$ less than the peak bin frequency value stated in a). The fitted peak does not incorporate the negative skew so the population mean value is more likely to be towards the peak bin value. The Svalbard distribution's fitted Gaussian mean value in b) is -11ms$^{-1}$ with the same error of $\pm$0.2ms$^{-1}$ (small compared to the standard deviation). This Gaussian mean values is only 1ms$^{-1}$ away from the peak bin value of -12ms$^{-1}$. This alludes to how well the Gaussian fits the Svalbard distribution if the positive tail is ignored; the positive tail is the only real source of kurtosis. The standard deviation of the fits are 16$\pm$0.2ms$^{-1}$ for Svalbard and 24$\pm$0.2ms$^{-1}$ for Kiruna.

According to the distributions, it seems Kiruna's winds are generally larger magnitude than Svalbard's (wider standard deviation); complying with Kiruna's relative proximity to the auroral oval. The bulk of Svalbard's winds are confined to a smaller range of values which are mostly downward. Conversely, nearly half of Kiruna's wind events are downward and half upward. However, both sites have equal probability of observing extreme upward wind events >70ms$^{-1}$ (see how the positive tails overlap in a)), this apparently not complying with the two site's relative proximity to the auroral oval. Kiruna would be expected to have proportionately more extreme vertical wind events subject to the greater number of times the auroral oval is expected to be overhead.

Kiruna has more extreme negative wind events than Svalbard, shown by its distribution's negative tail extending towards $\sim$-150ms$^{-1}$ (symmetric to its positive tail) and Svalbard's extending towards $\sim$-100ms$^{-1}$ (not symmetric to its positive tail). This complies with the expectation that the magnitude of winds is smaller for a polar cap site compared to an auroral oval one (as the relative standard deviations also suggest). However, the positive tail, as we have seen in the previous paragraph, does not comply with this expectation.

Svalbard's distribution is surprising. Its fat positive tail is not matched with a corresponding negative tail. This is expected if the extreme upward winds were attributed to a heat source. A symmetric negative tail would plausibly represent the adjacent convective subsidence of the atmosphere. It is unusual to have a distribution off-centred, particularly for vertical wind which is usually assumed to average to zero (Aruliah et al, 1994)

### 4.3.2 Magnetic Local Time Split Distributions

In section 4.1 we have seen that Svalbard and Kiruna are statistically expected to be underneath the auroral oval at different times of the day. Therefore, by splitting the distributions according to time of day, this will show the vertical wind's dependence on proximity to the auroral oval. It may also aid the understanding of the extreme vertical winds evident in Figure 4.9b). The data are split by magnetic time sectors (MTS), i.e., into four 6 hour time bins centred about magnetic noon, dusk, midnight and dawn.

Figure 4.10 displays the Svalbard (sub-plot a)) and Kiruna (sub-plot b)) data split into four 6 hour MTSs. The noon sector is in blue, the dusk sector is in teal, the midnight sector in gold and the dawn sector in pink as labelled in the legend to the left of the distributions. Statistics for each sector are listed in rows in line with the legend. These are peak bin value, the mean of the data $\mu$, the standard deviation of the data $\sigma$, the kurtosis of the distribution $\kappa$ and the skewness of the distribution $\gamma$. There is insufficient data in Kiruna's midday sector to create a distribution therefore the Kiruna distributions in figure 4.9 are limited to 15-06MLT. By comparing each MTS, we eliminate bias. To ease distribution cross-comparisons, each has been normalised to the maximum number of data points per bin in each distribution. This removes sample size information. However, since the smallest distribution (dawn) has a peak frequency of 220 data points and is a smoothly varying shape, we can assume that all distributions have a large enough sample to make inferences from.

Considering Svalbard first, the most prominent feature of this plot is the protrusion of the midnight distribution above the others in the positive vertical wind space. The noon, dusk and dawn sectors' distributions are very similar to each other. Their peak positions are -11$\pm$2 ms$^{-1}$, their widths are identical ($\sigma = \pm 15 \mathrm{ms}^{-1}$), they are all mildly positively skewed and all have large kurtosis values ($\kappa > 9$), due to their elongated positive tails. The midnight distribution has a lower kurtosis ($\kappa = 6.4$) due to the positive tail broadening out causing a larger fraction of the data to be closer to the mean. This also results in the standard deviation being much larger ($\sigma = \pm 25 \mathrm{ms}^{-1}$) than for the other MTSs. The Svalbard midnight distribution is visually the most assymetric however this is not reflected in the skewness value of 1.97, which shows the dusk distribution to be more skewed ($\gamma$=2.69). The skewness value is calculated as the mean cubed difference between the data and the data mean divided by the standard deviation. Due to Svalbard's bulky

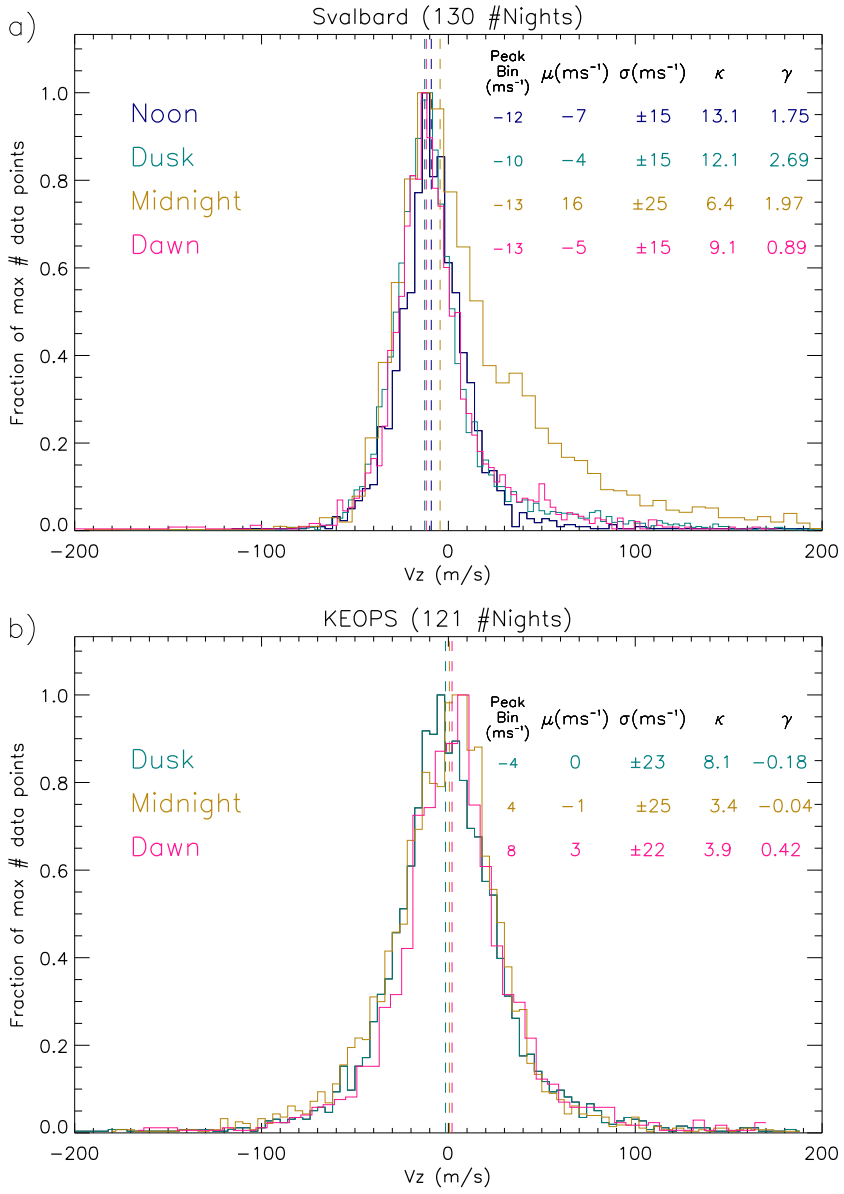


Figure 4.10: Frequency distribution of vertical wind data for Svalbard (top plot) and Kiruna (bottom plot) from 2002-2009 November-January (all accompanied by ASC data). The data are further split by six-hour wide magnetic time sector centred on magnetic noon (blue), dusk (green), midnight (yellow) and dawn (pink), which are over-plotted on one another as separate frequency distributions for each plot. An attempt to fit a Gaussian has been done and labelled to the side are: the peak bin value of the distribution, the mean of the data $\mu$, the standard deviation of the fit $\sigma$, the kurtosis of the distribution $\kappa$ and the skewness of the distribution $\gamma$. These are colour-coded like the distributions for each magnetic time-sector. There is no noon distribution for the Kiruna FPI does not take data from 9-15MLT as it is too light (too low latitude).

positive tail the mean of the data (16ms$^{-1}$) is positive and much larger than the peak value (-13ms$^{-1}$). This, together with the relatively large standard deviation value, means the skewness of the distribution is not as large as it seems when visually comparing the assymmetry between the different MTS distributions.

Secondly, considering Kiruna, all three distributions are very similar; all leptokurtic, all have similar widths, all mean values are relatively close to zero (-1-3ms$^{-1}$) and all have negligible skews ($\gamma$=-0.18-0.42) compared to that of Svalbard's. It is interesting that there are roughly relative fractions of extreme vertical wind events occurring at each MTS, despite Kiruna being underneath the auroral oval most at midnight. This exhibits a very subtle effect that the auroral oval has on the vertical wind distributions.

In summary, all magnetic time sector distributions for both sites are leptokurtic. The Kiruna distribution is quite unremarkable and there is no obvious impact due the proximity of the auroral oval. It is the same situation for Svalbard's MTS distributions except for the midnight population which has a broad positive tail. This shows an excess of large to extreme vertical winds, even though this is the time sector where Svalbard is furthest from the auroral oval; a very intriguing phenomenon is occurring. The thermospheric dynamics and energetics in the polar cap needs to be investigated in the midnight sector in order to determine this effect. Although it looks like this phenomenon is occurring a residual amount in the other time sectors also as can be seen in the elongated tails.

### 4.3.3 Conclusion

To conclude, disparities have been found between the Svalbard and Kiruna vertical wind data and their relationships with the auroral oval. The Kiruna data, which is more representative of auroral activity, shows on average smaller vertical winds with a subtle relationship to the auroral oval. The Svalbard data however, has revealed an excess of extreme (and anomalous >175ms$^{-1}$, i.e., 5-sigma away from the mean) positive vertical winds around midnight. These data explain the pronounced trend (maximum at midnight) seen in the average variation with MLT shown in figure 4.7. This trend is shown to be more enhanced under quiet conditions. From section 4.1 it is established that Svalbard is furthest away from the auroral oval in the midnight sector, particularly under quiet conditions. This indicates a counter-intuitive relationship where the vertical winds are larger, the further away from the auroral oval Svalbard is. This is opposite to Kiruna's relationship which shows more positive vertical winds when closer to the auroral oval at

midnight. These positive winds are very small (<5ms$^{-1}$) in comparison to the 20-40ms$^{-1}$ at Svalbard.

Considering the data in histogram form revealed that all populations (both sites and all MTSs) are leptokurtic. This shows an excess of data distributed far away from the mean. This is an attribute of very energetic localised physics. The neutral medium is so viscous and the thermosphere so convectively stable, it is usually expected that extreme vertical winds could not be generated and the histograms produced would comply with the normal Gaussian distributions. Perhaps the smaller leptokurtic tails of the Kiruna distribution (3.4< $\kappa$ <8.1) could be explained by auroral heating. This is not an option for the more leptokurtic Svalbard distributions (6.4< $\kappa$ <13.1) which represent behaviour poleward of the auroral oval.

Svalbard's distributions' peak locations are negative, roughly -10ms$^{-1}$, whereas Kiruna's are always about zero. Svalbard's are contrary to traditional perceptions of thermospheric vertical wind, which are expected to average to zero. Furthermore, the asymmetry in Svalbard's distributions do not seem natural. Suppose there is a strong large-scale heat source of energy in the polar cap driving these extreme upwellings. Nearby, downwellings of equal or similar magnitude would be expected and these are not evident on the negative side of the distributions. Kiruna's distributions, which do show some upwellings, probably driven by the auroral oval, also show downwellings on the negative side of the distribution of a similar magnitude, as expected.

There is plausibly some other physical processes that occurs in the polar cap (Svalbard) that is distinct from the auroral oval processes (Kiruna) and produces these negatively offset distributions, with highly elongated positive tails and relatively large kurtosis values. At midnight this excess of extreme vertical winds is largest. However, there are also residual levels of extreme vertical winds in the other MTSs. Svalbard's heavy positive tail is now termed a 'black swan' tail (Taleb, 2007). This alludes to the nature of the events that make them up; events that are considered impossible but do, in actuality, have a finite chance of occurrence. Taleb (2007) defines them using three criteria: firstly they are statistically unusual (or surprise results), secondly they have a large impact on the physics (if extreme vertical winds are commonplace then hydrostatic equilibrium must be frequently violated) and thirdly, upon being found, they may be rationalised with hindsight. A related analysis is presented in the following investigation in chapter 5. The leptokurtic distribution shape is typical of black swan distributions.

We should be cautious when characterising the vertical winds from these distributions when we do not know the source of the leptokurticity and Svalbard's heavy positive tail. Although the leptokurticity and heavy positive tail are unlikely to affect the shape of the bulk of the distribution (close to the mean). Therefore, it is reasonable to compare the widths of the sites' distributions. Kiruna's standard deviation is wider than Svalbard's which is indicative of the site spending more time under the auroral oval, subject to the process of energy deposition enhancing wind variability. This finding gives further evidence of Svalbard lying outside the auroral oval more often than Kiruna which adds to the surprise of the black swan tail.

In conclusion, Kiruna's distributions seem to be interpretable and comply with the known physics of the auroral region, whereas those of Svalbard do not. In the next section we investigate the Svalbard distribution's black swan tail and the mechanism by which it broadens. The possibility of the black swan events being spurious is considered which would then pose doubts upon previous authors' findings of large vertical winds at midnight in the polar cap (Spencer et al, 1982; Rees, 1984a; Innis et al, 1996, 2001; Greet et al, 2002).

This page was intentionally left blank

# Chapter 5

# Svalbard's Black Swan Tail vs. Spectral Contamination

## 5.1 Introduction

The Svalbard vertical wind distributions all exhibit an excess of extreme upward winds. These could potentially be black swan events which are usually identified from the tails of heavily leptokurtic distributions. There are three criteria by which these events may be classified as black swan events (Taleb et al, 2007). The first two are satisfied for the present case: i) a surprise result (i.e., an unexpectedly large excess of extreme positive vertical winds in the polar cap), and ii) has a large impact on the physics (i.e., likely to be violations of hydrostatic equilibrium and altitudinal displacement of chemical species altering thermospheric chemistry). The third, rationalising with hindsight, is the aim within this chapter. The possibility that the tail is due to spurious data, such as a spectral contaminants or an instrumental effect, is explored. If found to be erroneous, the tail events cannot be considered black swan events and measures to clean the data will be explored.

Examining some black swan event case studies may provide us with some insight into the relevent physics. Figure 5.1 displays sky parameter line plots for six case studies

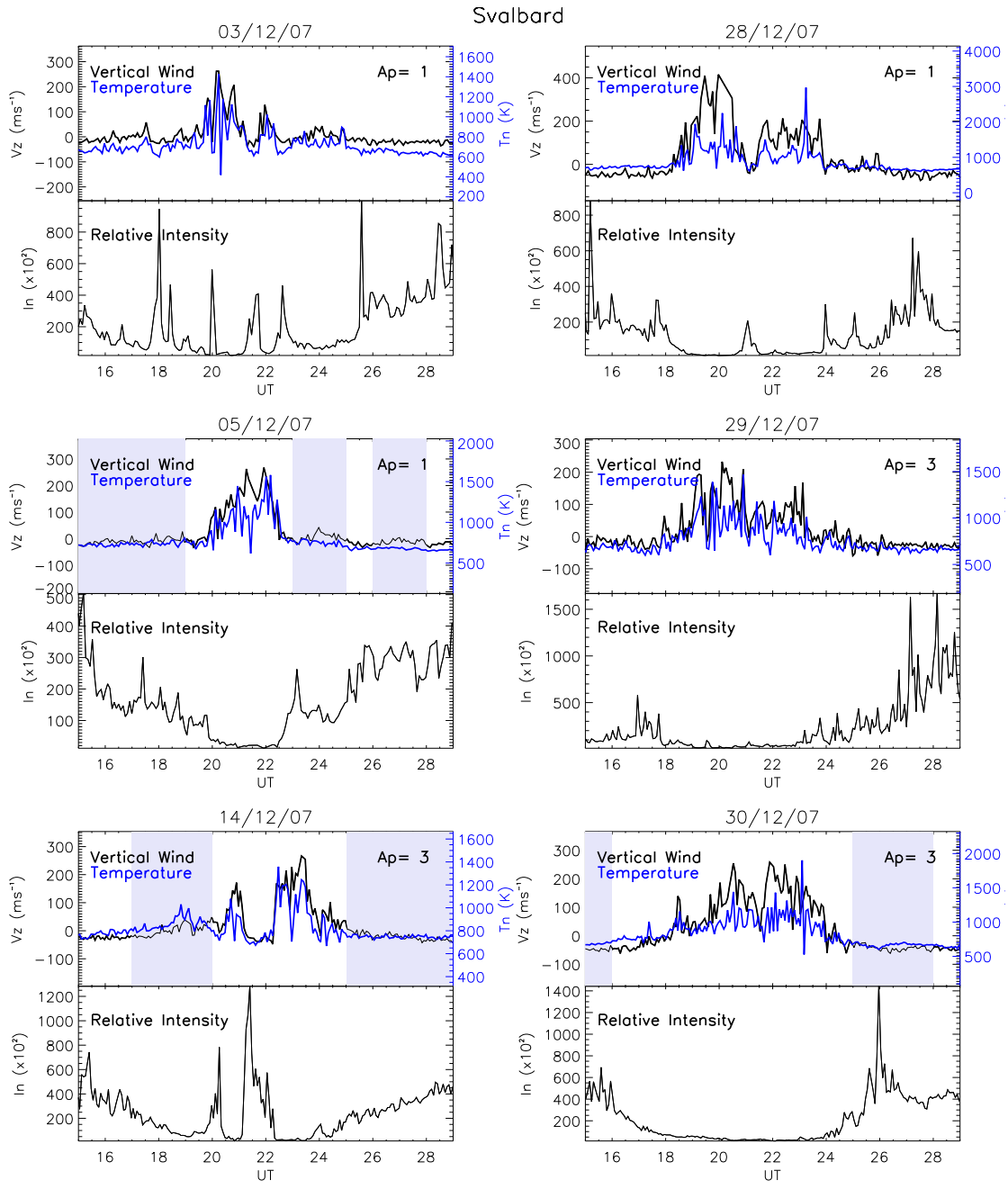


Figure 5.1: Svalbard sky parameter data with UT for 6 nights in December 2007 exhibiting black swan events, i.e, vertical winds >100ms$^{-1}$. There are two sub-plots for each night (date labelled in plot titles). Vertical wind (black) and neutral temperature (blue) are plotted in the top sub-plots and intensity is plotted in the bottom sub-plots.

from December 2007 where the upward wind magnitude exceeds 200ms$^{-1}$. There are two plots for each night, the top displays vertical wind (black) and temperature (blue) and the bottom displays the relative 630nm emission intensity variation with UT. The background shading represents times of cloud cover during which the data are unreliable. There are many more examples of nights like this throughout the years 2002-2009. However, these six have been chosen as they are accompanied by the most supplementary data-sets, e.g., SCANDI and EISCAT.

These nights all show an intriguing phenomenon. The upwellings all occur around magnetic midnight (∼21UT), for a period of 2-3 hours and are coincident with neutral temperature enhancements (data measured by FPIs of Doppler broadening of 630nm emission) of approximately 200K and significant drops in emission intensity. Furthermore, these events occur on nights of very low geomagnetic activity ($0<A_P<3$). Every upwelling reaches at least 200ms$^{-1}$ and for 28th December (top right plot) ∼400ms$^{-1}$ is the maximum wind speed. Due to every upwelling being closely accompanied by such large temperature enhancements, a causal relationship is indicated; there seems to be a heat source driving the winds. These unprecedented magnitudes and the persistent nature of the winds imply a very energetic heat source, typical of the auroral oval.

The centralisation of the upwellings about magnetic midnight, at first appearance, indicates that the auroral oval may be the driving heat source as the midnight auroral region has the most intense energy deposition (from hard particle precipitation and the auroral electrojet current's Ohmic heating). However, from section 4.1, we know that Svalbard is always poleward of this midnight auroral region, particularly on geomagnetically quiet nights like these. This is corroborated in figure 5.1 by the emission intensities being relatively low throughout each upwelling period. Evidently, the auroral oval is not overhead.

For such extreme vertical winds as these at times of such low emission intensity, it is necessary to study the raw spectral profiles from which the Doppler shifts were derived. Figure 5.2 shows spectral profiles from the case study nights used in figure 5.1. The derived vertical wind values are labelled in the top left hand corner of the plot and the signal-to-noise ratio (SNR) is labelled on the right hand side of each plots' header. The SNR is defined as the power spectrum's fundamental component divided by the average value of it's high-frequency, noise dominated components (please see section 2.6.3 for more detail). Generally, for these spectra, a 'good' value is thought to be roughly >2500.

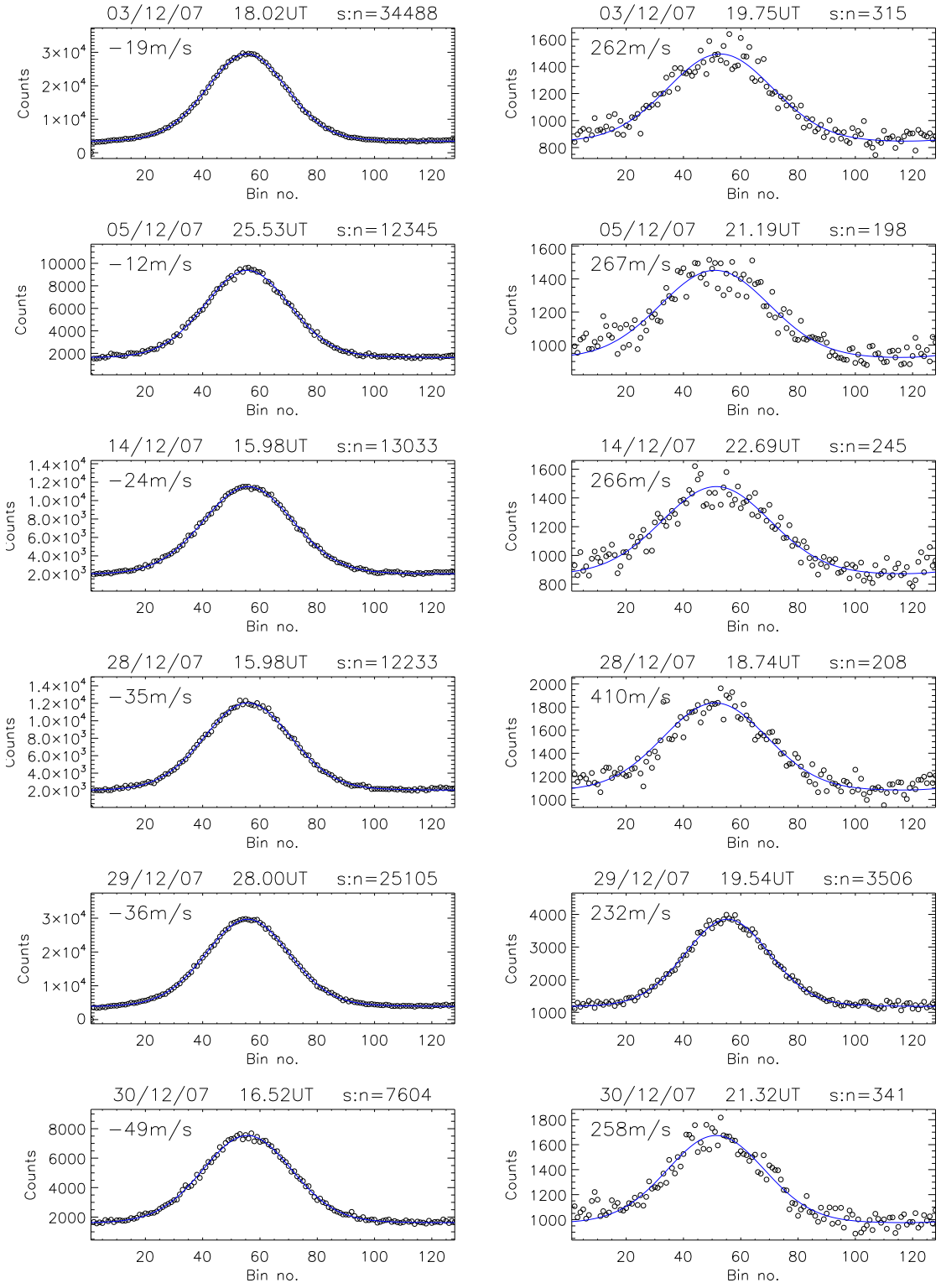


Figure 5.2: Spectral profiles from the six Svalbard upwelling case studies from December 2007 used in figure 5.1. Each row is for a different night. There are two spectral profiles displayed from each night of data, i.e., 2 columns of plots. The left-hand-side plots show examples of high signal-to-noise ratio (7604-34488) spectra from the night and the right-hand-side plots show the spectra which give the maximum vertical wind values from each night. These also show low signal-to-noise ratios (198-3506).

Within figure 5.2, each row of plots displays data from a different night of the case study. The plots on the left hand side show a randomly chosen example of a spectral profile with a strong SNR from the same night of data. These good quality profiles are used as comparison plots for the profiles to their right. The right hand plots display the spectral profiles representing the maximum velocity value from each upwelling displayed in figure 5.1. These all happen to be poor quality data. Five out of the six spectral profiles have extremely poor SNR (for FPI spectrograms) ranging from 198 to 341. The only one of these spectral profiles with an acceptable value, at 19:54UT on 29th December, only marginally exceeds the threshold, with a value of 3504. These six upwelling spectra also exhibit a dramatic reduction in intensity when compared to the good quality spectra to their left. There is no obvious additional spectral line evident on the upwellings' spectra, although the noise could be obscuring a low intensity line. Nevertheless, it is problematic to have confidence in the accuracy of these wind derivations when their corresponding spectra are not robust. For the left-hand-side spectra there is much more confidence in the derived values. The derived velocity values for the left-hand spectra are all negative. Furthermore in figure 5.1, it is evident that at the times before and after the upwellings the vertical wind is mostly negative. This is odd; vertical wind fluctuations are expected to hover around zero (Aruliah et al, 1994). This could be the same phenomenon as the distributions peaking $\sim$-10ms$^{-1}$ in chapter 4. These negative values could indicate a spurious zero velocity baseline being calculated due to spurious positive wind values (from the upwellings) being included in the average. The inclusion of the spurious positive values means the the average raw vertical wind (i.e., the baseline) would be much more positive than it actually is. When subtracted from the vertical wind signal (i.e., baseline correction procedure (see section 2.6.4 for a more detailed explanation)), it is over-corrected so that the 'good' data are more negative than they are supposed to be in reality had the spurious data not been included in the baseline calculation.

The example spectra shown in figure 5.2 exhibit a positive correlation between noise and vertical wind velocity that is found to be evident throughout the whole dataset. The case-study parameter plots in figure 5.1 clearly showcase the correlation between the vertical wind and temperature which are anti-correlated with intensity. It is necessary to find a plausible mechanism for this phenomenon while, at the same time, exploring the possibility of the upwellings being spurious due to some systematic spectral distortion about midnight. Some potential explanations for such phenomena could be:

1. Real upwelling events driven by some high latitude heat source other than the auroral oval

2. An instrumental effect such as vignetting, cosmic ray events or stars, which distort the spectra

3. Spectral contamination, from species such as hydroxyl or hot oxygen, which distort the spectra

These three explanations are discussed consecutively in the three following sections 5.2 - 5.5.

## 5.2 Real Upwelling Events

In this section the possibility that the extreme upwellings are real is explored. This is done by considering the principle of convective stability; considering the horizontal wind behaviour at the time of the upwellings; and also by considering the distribution of vertical wind and temperature data at low signal-to-noise ratio.

### 5.2.1 Convective stability

Real upwelling events are produced by a heat source causing a parcel of gas to expand, rise and adiabatically cool until it reaches its new height of hydrostatic equilibrium. Often parcels convectively overshoot due to momentum and are brought back to equilibrium via a restoring force. As discussed in section 1.2.1 the thermosphere hosts the strongest restoring force due to its strong positive temperature gradient or ambient lapse rate. This means it is the most convectively stable region of the atmosphere. Therefore vast amounts of energy are required to drive an upward wind in the thermosphere, especially for periods of two to three hours. To clarify, Price et al (1995) estimate $10\text{mWm}^{-2}$ is required to drive a wind of $140\text{ms}^{-1}$ at 240km altitude. Innis et al (1999) use this value to derive a power input of $6\times10^{9}$W required to drive an upwelling of this magnitude over a typical upwelling area of $6\times10^{11}\text{m}^2$. For context, Roble et al (1987) estimated a global Joule heating power of $7\times10^{10}$W and a global particle precipitation power of $10\times10^{10}$W. This shows the power input required to drive a large upwelling is a non-negligible fraction of the global energy budget.

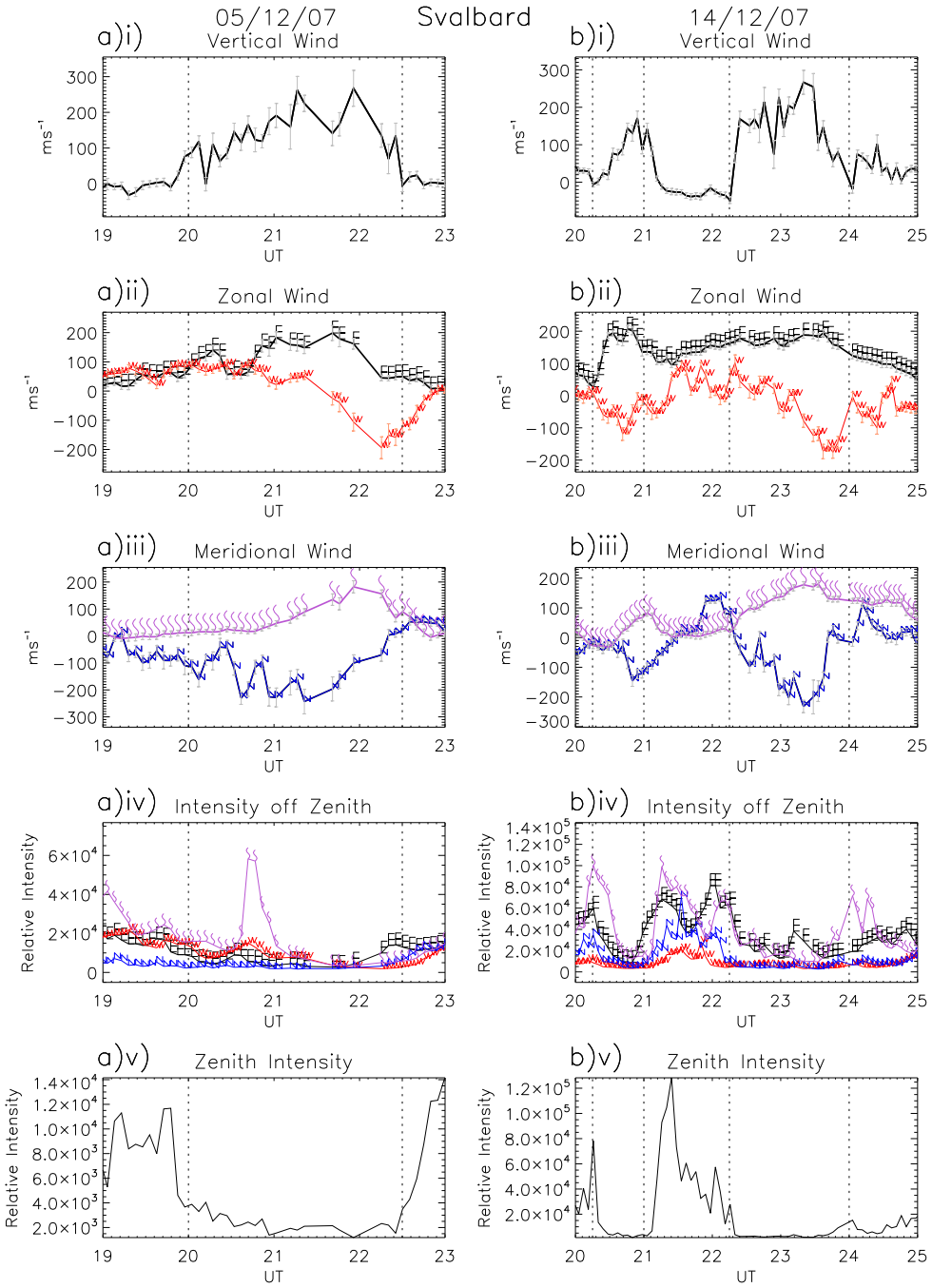


Figure 5.3: Line plots of the vertical and horizontal winds at 240km (measured by the Svalbard FPI 630nm line) with their corresponding intensities for two of the extreme vertical wind events. The 05/12/07 is displayed in sub-plots a) and 14/12/07 is displayed in sub-plots b). For each night, the first row of sub-plots i) show the vertical wind (zenith look direction) and the second row, sub-plots ii), show the zonal wind (east and west look directions, +ve east). East is coloured black with individual data points labelled as 'E' and west labelled red with 'W'. The third row, sub-plots iii), show the meridional wind (north and south look directions, +ve south). South is coloured purple with individual data points labelled 'S' and north labelled blue 'N'. The fourth row, sub-plots iv), show the intensity of the emissions for off-zenith look directions with the same colour code as in ii) and iii). The last row, sub-plots v), show the zenith look direction emission intensity. Overlaid on all plots are the boundaries in which the upwelling occurs

### 5.2.2 Horizontal Wind Enhancement

Continuity of mass means that the horizontal flux of gas is closed by the vertical flux. Thus the presence of a divergent horizontal wind field can be used as evidence of a vertical wind. However, due to light scatter, it is possible that wind measurements may be contaminated by photons from different look directions. This can cause a false detection for either the horizontal divergence or the vertical wind. Either an off-zenith photon, representing a real divergent horizontal wind, scatters into the zenith FOV producing a false upward wind; or zenith photons, representing a real extreme upward wind, scatter into off-zenith directions producing a false divergent wind. Both scenarios show an apparent upwelling-divergence relationship seemingly compliant with mass continuity.

On the nights of our upwellings the ASC has been checked to visually confirm clear skies so it was initially assumed that scatter could not occur. However, a thin haze or aerosols present in the sky, which are invisible to the ASC, are capable of scattering photons if the conditions are right. If there is a strong signal from all regions of the sky then the few scattered photons' signal will be insignificant within the resultant spectra. However, if there is a strong intensity gradient across the sky, the proportionately larger number of scattered photons from the brighter region of the sky may have a strong enough signal to be visible above the spectral signal of the dimmer region of the sky and thus bias the resultant spectra, manifesting as a spurious wind.


Figure 5.4: Example ASC image of an auroral arc to the south-east on 28th December 2007 (a case study night).

Evident in every December 2007 case study, at all times of upwelling, are large divergent horizontal winds. Figure 5.3 shows two examples out of the case studies to showcase this phenomenon. The data from the 5th December 2007 is displayed on the left hand side plots labelled 'a)' and the 14th December 2007 is displayed on the right labelled 'b)'. Different parameters are plotted with UT in each row. In row i), the vertical wind is plotted, ii) the zonal wind is plotted with the black 'E' line representing the eastward look direction and the red 'W' line representing the westward look direction, iii) the meridional wind is plotted with the purple 'S' line representing the southward look direction and the blue 'N' line representing the northward look direction, iv) the intensities of off-zenith look directions labelled the same as the previous two rows and v) represents the zenith intensity. Each upwelling of interest is roughly sectioned off by two dashed vertical lines through all plots to guide the readers eye to the times of interest. For a) the upwelling time is 20:00-22:30UT and for b) there are two upwelling times 20:15-21:00UT and 22:15-24:00UT. These sections show an increase in the vertical wind by at least 200ms$^{-1}$ and show a dramatic decrease in zenith intensity. These upwellings are accompanied by divergences in the horizontal wind with wind perturbations of the order of 200-300ms$^{-1}$ away from the zenith in all meridional and zonal directions, i.e., seemingly large scale divergent flows. The divergences occur when the corresponding off-zenith intensities are low. Sometimes the south and occasionally the east look directions have intensity fluctuations due to the auroral oval statistically lying to the geographic south-east of Svalbard (n.b., the FPIs' look directions are in a geographic reference frame). In the ASC these can be spotted as auroral arcs along the south-east edge of the image (see figure 5.4). When there is a fluctuation in the intensities in the south or east look directions, the corresponding wind ceases to contribute to the divergence (strongly positive outward flow) and reverts back to the ambient flow velocity (which is roughly hovering around $\pm$100ms$^{-1}$ at times of small vertical wind). The strongly divergent behaviour only occurs when all look directions' signal intensities are low.

**Upwelling-Divergence Correlation Mechanism**

The fact that the upwelling is always accompanied by a divergent wind could be due to four mechanisms:

1. Zenith photons (with large Doppler shift) are scattering into the off-zenith field of view, i.e., real upwelling, false horizontal wind

2. Off-zenith photons scattering into the zenith field of view, i.e., real horizontal winds, false upwelling

3. Mass continuity is being satisfied (see work on Burnside chapter 6), i.e., real upwelling and real horizontal winds

4. A different emitting species is contaminating the off-zenith spectra as well as the zenith, false upwelling and false horizontal winds

The relative intensity of the zenith and off-zenith signals are useful parameters to test the plausibility of these different mechanisms as follows.

**Mechanism 1: Zenith photons scattered into Off-zenith**

Zenith photons being scattered into the off-zenith FOV is initially believable in figure 5.3 due to a divergent horizontal wind accompanying an upward wind. A zenith upward wind is measured as positively Doppler-shifted photons, i.e., red-shifted. If these are scattered into off-zenith look directions, the off-zenith look directions' winds will appear to have positive Doppler shifts. These will manifest as a wind field flowing outward from the FOV, i.e., a divergent wind, complying with our data. However, due to the weakness of the zenith intensity signal, the likelihood of zenith photons being scattered is low as there are so few photons available. The first suggested mechanism is unlikely.

**Mechanism 2: Off-zenith photons scattered into Zenith**

Off-zenith photons being scattered into the zenith FOV is unlikely as, generally, the off-zenith intensities are so low there are few photons to undergo scattering. However, occasionally there are large auroral intensity fluctuations to the south and the east providing an abundance of photons to be potentially scattered. This mechanism would explain the large vertical wind magnitude as the contaminating horizontal wind magnitudes (from off-zenith observations) are relatively large. The first problem with this mechanism is that the south and east intensity fluctuations do not occur for the duration of the upwelling. The second problem is that it is suspicious that the off-zenith photons are always red-shifted, i.e., a persistent divergence, since variability is expected in the midnight polar cap horizontal winds.

**Mechanism 3: True Mass Continuity**

The mechanism of mass continuity explains the horizontal wind behaviour well in that upwellings are expected to be accompanied by horizontally divergent flow through the Burnside relation. The mathematical derivation of this result can be found in section 6.2. The main problem with this hypothesis is that the intensity signals are so low at the times of this correlation, the data are unreliable. Furthermore, when the signal becomes stronger and the data thus more reliable, the complementary behaviour between the two components of the wind (i.e., horizontal flow and vertical upwelling) is not so pronounced.

**Mechanism 4: Spectral Contamination**

The potential for another emitting species to contaminate the spectra also may explain the data. If this contaminating line is always showing up in a similar spectral channel it will always show a similar Doppler shift, no matter the look direction, hence could explain the horizontal wind fluctuation magnitudes matching the upwelling magnitudes. This condition also complies with the large wind velocities being measured when the signal intensity is low; this is when a spectral contaminant may especially become visible above the main spectral line. There is nothing to suggest that this mechanism does not explain the data. The phenomena leading to the suggestions of the three previous mechanisms (1-3) can also be explained by this mechanism. Therefore, this is the most plausible mechanism of the four possibilities so far.

### 5.2.3 Distribution Upturn at Low Signal-to-Noise

Considering the sky-parameter data distributions with different levels of SNR (see section 2.6.3 for the description of the calculation) can serve as a good diagnostic measure for peculiarities in the dataset. Ideally, SNR is the most direct measure of the quality of the data. However, SNR is not available for the whole Svalbard and Kiruna dataset because two different data-fitting procedures have been used, and only one outputs the SNR (n.b., both are very similar procedures; the one used up to this point uses laser profiles to estimate the neutral temperature and outputs SNR, however, so far, only a subset of the entire FPI database is analysed this way, therefore a small statistical sample; and the new procedure mentioned does not calculate temperatures or output SNR, however, there is a complete analysed FPI database, therefore a large statistical sample). For the larger

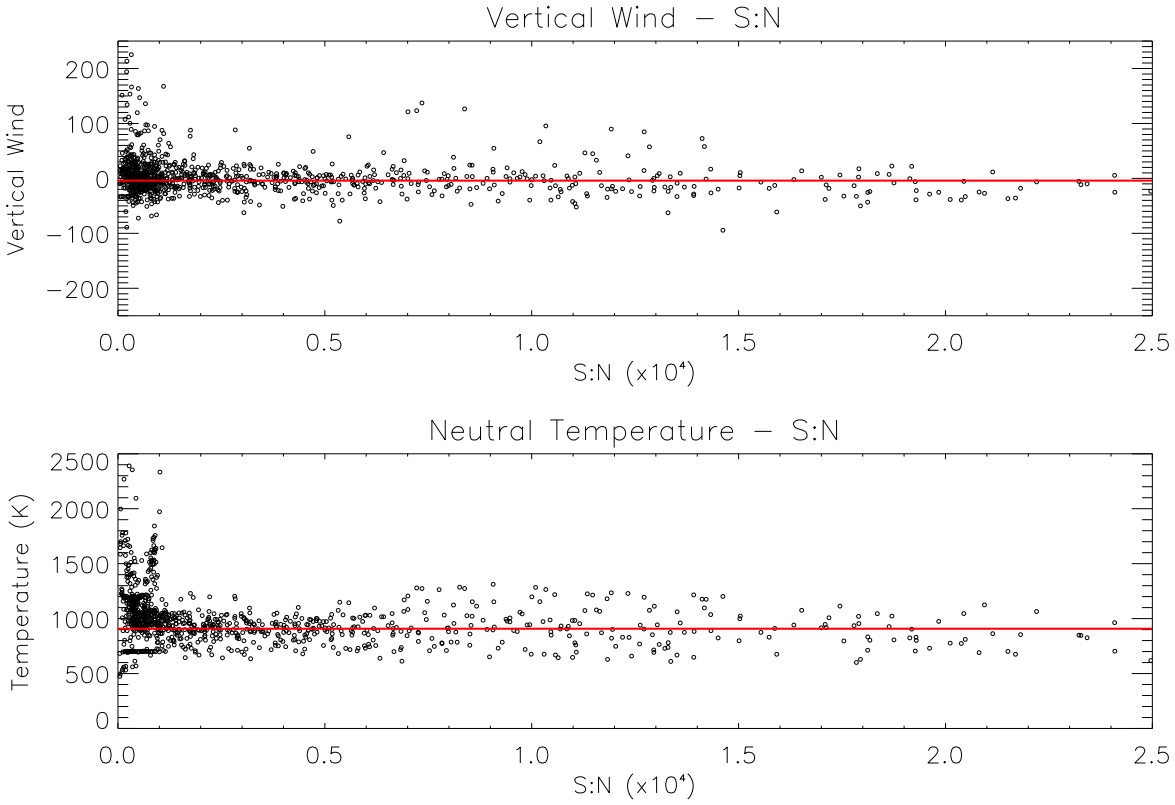


Figure 5.5: Svalbard red line FPI data-set from 2007-2009 including temperatures and signal to noise ratio. The first plot shows the behaviour of vertical wind with signal-to-noise ratio and the second plot shows the behaviour of neutral temperature with signal-to-noise ratio. The red lines show the mean value of the parameters calculated from the data $>0.2\times10^4$ signal-to-noise ratio, i.e., the symmetric data.

dataset, intensity can be used as a proxy for SNR (SNR and intensity are very strongly correlated parameters, shown following). There is SNR data for a few of Svalbard's seasons and these are displayed in figure 5.5 as two plots. The first one displays all data points of vertical wind against the SNR and the second plot shows the neutral temperature with SNR.

In the first and second plot, the main point to note is the asymmetry of the distributions at low SNR. It is expected that at low SNRs the uncertainty will be larger for these measurements. The asymmetry is unexpected though; there is no clear reason why a low SNR upward wind (or larger temperature) would have larger uncertainty than a low SNR downward wind (or smaller temperature). The red line shows the mean value of the parameters calculated from the data $>0.2\times10^4$ SNR, i.e., the symmetric data, ignoring the low SNR assymetrical artefact. It shows an estimate of the average line through the data points if the low SNR artefact is considered spurious and removed. At high values of

SNR, the points are evenly distributed about the line. Tracing these line towards the low SNR part of each plot, there is a clear upturn to the distributions on the positive (or more positive) side. There is no aeronomic reason for the vertical wind and the temperature to increase at low intensities.

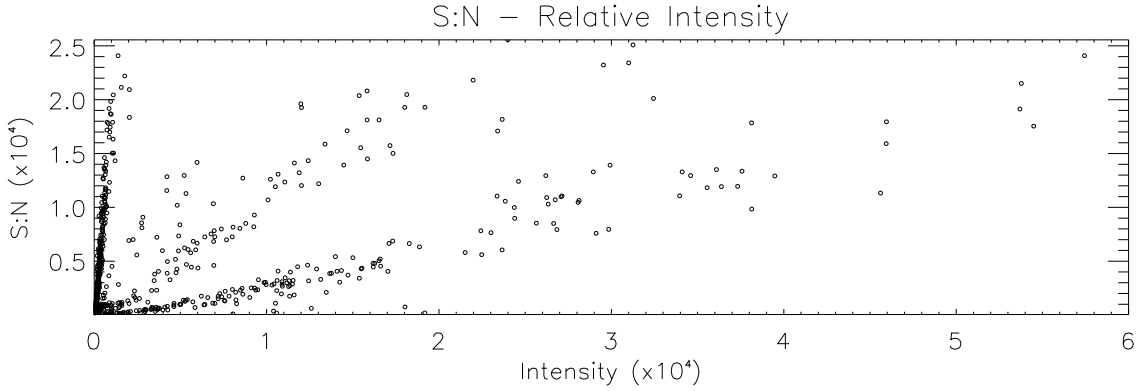


Figure 5.6: Svalbard red line FPI data-set from 2007-2009 showing the the correlation between SNR and relative intensity. There are three branches due to the FPI changing its software set-up and drifting over the two year period (see figure 2.7).

To increase the statistical sample, intensity is used as a proxy for SNR, as there is a much larger already existing database of this quantity. Figure 5.6 uses the same 2007-2009 data as in figure 5.5 to show relationship between SNR and intensity. It is a scatter plot composed of three different branches of correlation with different gradients. Each of these represent different time periods when the FPI detector has experienced a different transmission intensity due to varying instrument sensitivities, mislaignment or output gain (see section 2.7). Nevertheless, each of these branches shows a proportional relationship with intensity increasing with SNR. This overall positive proportionality gives confidence in using the intensity of the signal as an an acceptable proxy parameter.

Figure 5.7 shows the vertical wind distribution with intensity for both Svalbard and Kiruna. For each day of data included, the intensity is normalised to the maximum intensity experienced in the day. This ensures the three different intensity categories (evident in the third plot of figure 5.5) are on a similar scale. The scales will not be exactly the same from day-to-day as the daily maximum intensity changes, however, the scales will be similar enough to reveal the approximate distribution shape of the vertical wind data at low intensities. The top plot shows the Svalbard distribution and the Kiruna one is displayed on the bottom for an auroral latitude comparison. For Svalbard the same upturn shape at low intensities, previously seen in figure 5.5, is clearly displayed here but

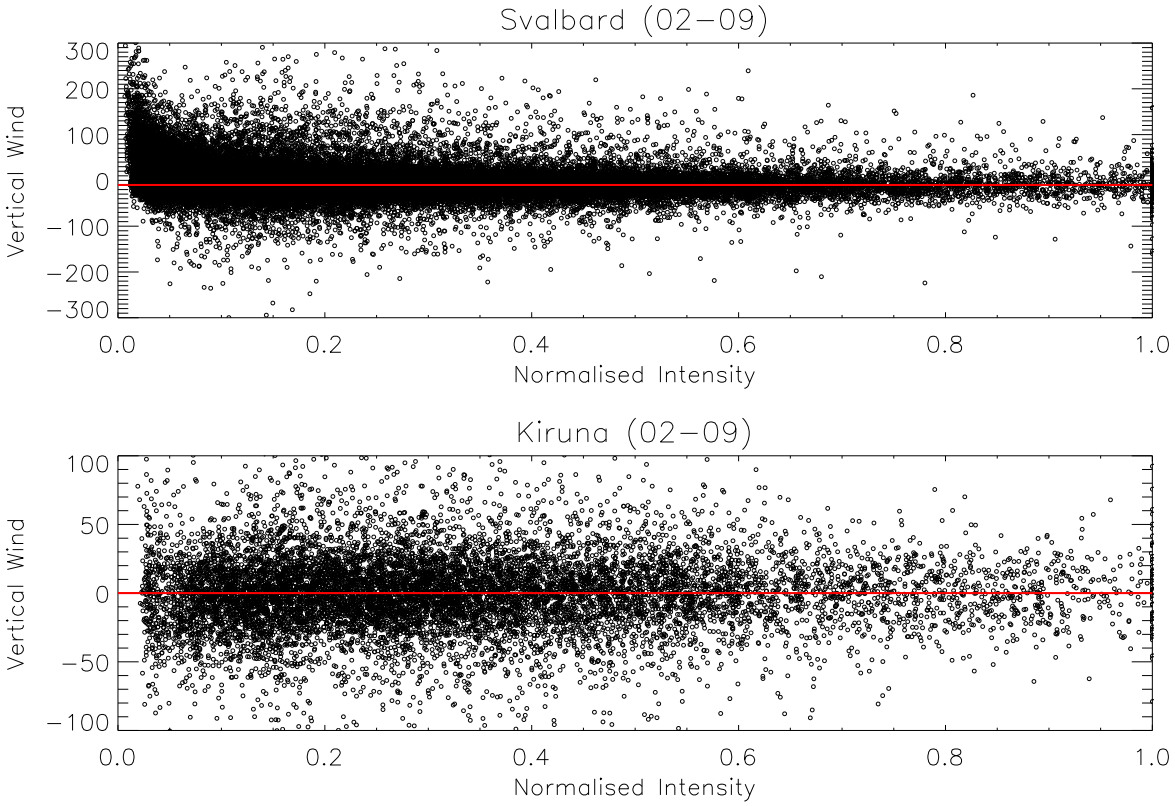


Figure 5.7: Svalbard and Kiruna red line FPI vertical wind data from 2002-2009 scatter-plotted with intensity which is normalised to the maximum value in the day. The red lines show the average parameter value according to high intensity data >0.4 noramlised intensity.

with a much higher density of data points, showing the effect with more significance. The Kiruna vertical wind distribution is symmetric. This shows the distribution expected from a site which is in close proximity to (or underneath) the auroral oval most of the time. The Svalbard distribution has a preferred direction of vertical wind at low intensity emissions. This cannot be related to any auroral forcings or this upturn feature would show up more strongly on the Kiruna distribution. It is proposed that a spectral contaminant becomes apparent at low intensities, and this will be discussed later.

### 5.2.4 Conclusion

In conclusion, it seems that the extreme midnight upwelling events are not real. The thermosphere, due to its large convective stability, resists large amplitude vertical winds. Therefore, any that do occur, require significant energy to maintain. The particle and Joule heating in the auroral oval are not powerful enough to provide this energy, particularly at

the geomagnetically quiet times of the case studies considered. Furthermore, if the auroral oval heating was driving the upwellings, Kiruna, which spends more time underneath the auroral oval, would exhibit the same phenomenon as Svalbard but more exaggerated. That is, Kiruna would experience more extreme upwellings and more pronounced upturn shapes would be present on the vertical wind and temperature distributions (with SNR or intensity) at low SNR (or intensity). Kiruna shows no upturn and, as we have seen in chapter 4, shows fewer extreme upwellings than Svalbard (see figure 4.8). This is further corroboration of the auroral oval not driving the vertical winds. It is highly unlikely that there is another undetected large scale heat source poleward of the auroral oval powerful enough to drive these extreme winds. This is because it would be a particularly powerful energy source, of the order $6\times10^9$W (Innis et al, 1999), which would certainly affect other polar MIT dynamics, and there have been no reportings of any such heat source in any previous space weather research.

The possibility of photons from different look directions scattering across the FOV is unlikely due to the low general levels of signal intensities. Furthermore, the upturns in the vertical wind and temperature parameter distributions (with SNR or intensity) can be interpreted as being due to some kind of spectral distortion. Therefore we conclude that the, previously termed, 'black swan events' are not true vertical winds. Instead, it seems they are the result of some kind of instrumentation effect or other contaminating species. These possibilities are addressed in sections 5.4 and 5.5. Svalbard's fat positive tail is now understood not to be a black swan tail as it is a result of a different kind of physics separate from thermospheric vertical winds.

## 5.3 Implications of Spurious Upwelling Events on Vertical Wind Distribution

The findings that the extreme upwellings evident in Svalbard's positive tail are not black swan events, but are spurious, has implications on the true vertical wind distribution. To correct the distribution, it is not as simple as removing the spurious data points from the distribution; the magnitudes of the other reliable data points are intertwined with the spurious data points through the baseline removal (see section 2.6.4). By making inferences about the resulting distribution, using the Kiruna distribution as a guide, a few slight adjustments can be made to the Svalbard distribution in order to superficially

correct it.

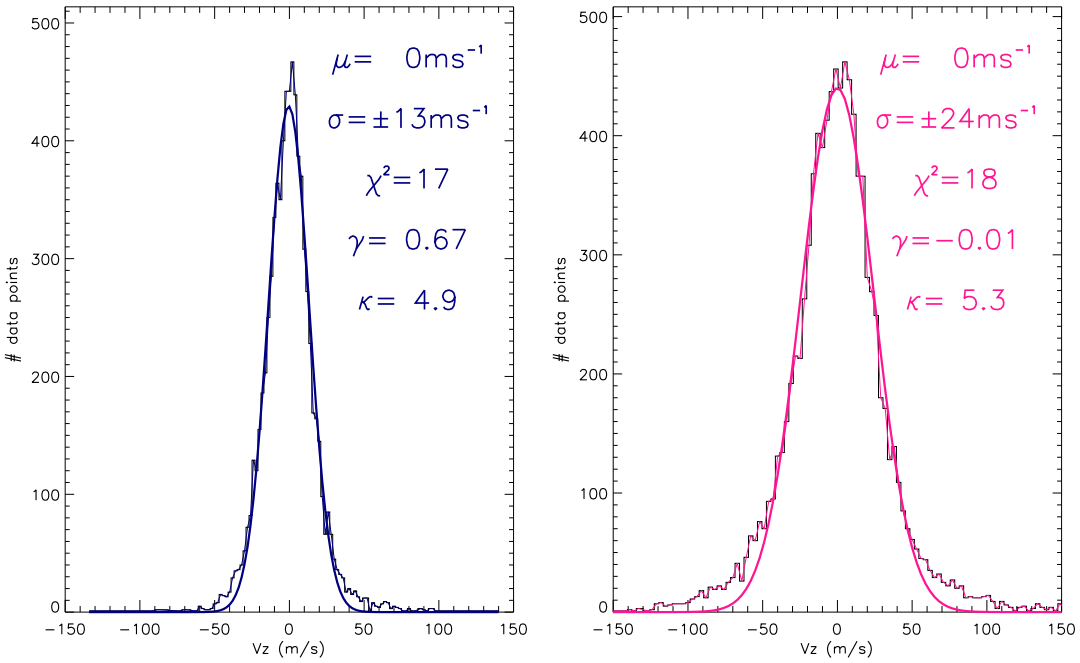


Figure 5.8: Svalbard (blue) and Kiruna (pink) red line FPI vertical wind data (2002-2009 Nov-Jan) in histogram format but with the spurious data removed from the Svalbard distribution

The FPI zero-Doppler shift baseline is calculated by assuming that, over an entire night, the vertical winds average out to be zero. Hence, the baseline is taken to be the average zenith wind over the night. By subtracting this, the data are recentred about zero and assumed corrected. However, if there are false extreme positive vertical winds (about midnight) included in the calculation, the average zenith wind will appear larger than it actually is and so, after subtracting this baseline, the remainder of the vertical winds will be erroneously shifted to the negative, i.e., apparently downwards. This explains the reason we have seen Svalbard's vertical wind distribution centred in the negative.

Kiruna's vertical winds are centred around zero, despite the fact that the site is subject to more energetic auroral forcings. A theory to describe the Svalbard vertical wind distribution shape (a negatively-centred positively-skewed distribution) is that it is indicative of being underneath the auroral oval subject to its heating. The auroral heating will drive large upwelling vertical winds (fat positive tail) and outside the auroral oval (most of the time for both sites) the vertical winds will subside over a larger area at a smaller magnitude (the negative centre). This theory is false as Kiruna's distribution is both symmetric and centred at zero. If Kiruna, which spends more time underneath the

auroral oval, is centred at zero and is symmetric, then Svalbard, which spends less time underneath the auroral oval (mostly poleward), is likely to be centred at zero and symmetric. Therefore, we hypothesise that Svalbard's actual vertical wind distribution will resemble the left-hand plot of figure 5.8. Two adjustments have been made to the original Svalbard vertical wind data in order to get this shape. Firstly, the low SNR FPI relative intensity data, $<4\times10^3$, has been excluded in order to remove any data which could contribute to the upturn feature data in figure 5.5, i.e., SNR$<2\times10^3$ (this corresponds to a maximum relative intensity of $<4\times10^3$ in figure 5.6). This removes the asymmetry. Secondly, the distribution is artificially shifted $+11\text{ms}^{-1}$ to the right. This centralises the distribution. This hypothesised distribution is now symmetrical and centralised just like the true Kiruna one, shown in the right panel of figure 5.8.

This hypothesis is tested at the end of this chapter in section 5.5.7 once the vertical wind data has been reanalysed after discarding the spurious data with new baseline calculations. Before this, a method must be developed to identify the spurious data which will be shown in section 5.5.

Considering the hypothesised Svalbard vertical wind histogram, it is proposed that the standard deviation of the data is likely to be approximately $\pm13\text{ms}^{-1}$. This is only $2\text{ms}^{-1}$ away from the original calculation of $\pm15\text{ms}^{-1}$ (see figure 4.9 b)). The small difference is likely to be due to the removal of the spurious data which is hypothesised to be from the fat positive tail leaving the width of the distribution essentially the same. We can infer upper bounds to vertical wind values by considering the negative tail of the distribution which represents the uncontaminated downward winds. If we shift the distribution $11\text{ms}^{-1}$ to the positive then the distribution tends to zero frequency at approximately $50\text{ms}^{-1}$. This is a good estimate of a particularly large vertical wind event; much more believable than the $200\text{ms}^{-1}$ events from our case studies (see figure 5.1). These revised estimates comply more with previous observations of high latitude vertical wind magnitudes being relatively small, i.e, $<40\text{ms}^{-1}$ (Smith ,1998; Conde and Dyson, 1994).

## 5.4 Instrumental

In this section the possibility that instrumental effects are causing the apparent extreme upwellings is considered. Other than transient spectral contaminants from the sky that the instrument cannot recognise and remove, e.g., cosmic rays and stars passing through

FOV, the only true instrumental effect that could potentially skew the spectral peak is vignetting, which will be the main topic of this subsection.

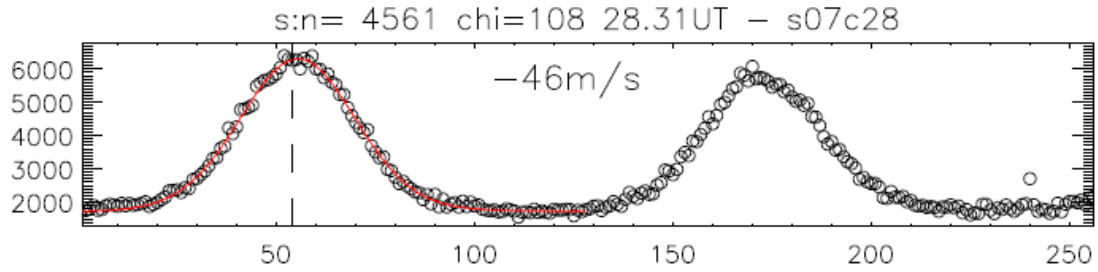


Figure 5.9: Example spectral profile with a signature of a cosmic ray event perturbing the intensity at bin 240.

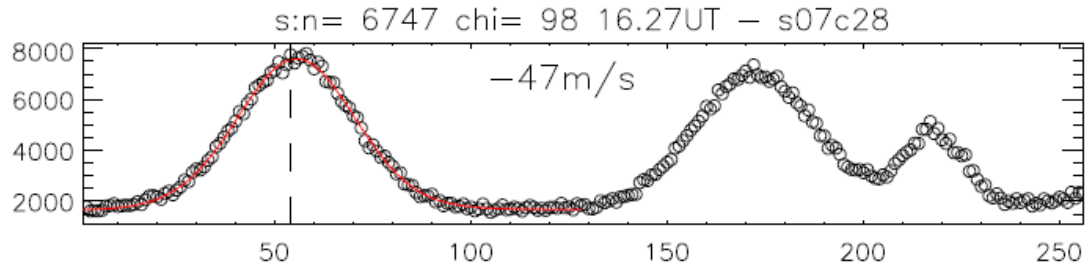


Figure 5.10: Example spectral profile with a signature of a star passing through the FPI field of view centred on bin 218.

An example of a cosmic ray event is shown in figure 5.9 which shows a spectral profile, i.e., relative intensity varying with spectral bin number. In bin number 240, there is a signature of a cosmic ray: a random intensity fluctuation above the smooth intensity variation of the spectral profile. There is no software currently in place to remove them; however, they are not too troublesome, only affecting one bin at a time. Cosmic ray events thus just add to the signal's noise rather than skewing the distribution. The signature of a star passing through the FOV is shown in figure 5.10. It manifests as an additional smaller peak on the spectral profile. In this case it is centred about bin 218. This can be distinguished from a contaminating line by the fact that it is not repeated on the second peak. This does not introduce much error as the chance of a star passing through the $1^o$ FPI FOV is very low and a transient phenomenon. Therefore these contaminants are not likely to be the source of the systematic shift in the fitted spectra we are seeing at midnight. Nevertheless we should be cautious of these effects when there is an isolated

data point fluctuating in vertical wind or temperature.

### 5.4.1 Vignetting

Vignetting is an instrumental effect which is always present. It refers to the image on the detector chip being brightest in the centre and reducing in intensity towards the edges. Vignetting is mainly due to off-axis rays not reaching the chip. Our instrument engineer, Ian McWhirter (1995), re-designed the optics optimally for our FPIs using ray-tracing techniques and considerably reduced the vignetting effect. However it can never be removed completely, due to various reasons:

- An amount of vignetting is inherent in the etalon itself because some off-axis rays hit the side of the etalon.
- Another source is from the interference filter's passband shifting with angle of incident light. This causes some off-axis rays to be outside the passband. To reduce this effect our filters are made with the passband slightly high to allow some off-axis rays to fall within the passband. However the passband of the filters drift due to ageing affects so this effect needs to be monitored.
- Another source of vignetting is uneven illumination. This is true of the laser scan taken before 2007, when the uniform calibrating panels were installed onto the instrument. Before this the laser was shone onto a diffusing screen in an attempt to spread the light out. Although more uniform, the central part of the image was still brighter. This is not an issue for the sky emission as this is an extended light source.
- Additional distortion of the peak can be caused by the x-y centre of the FPI image not being set correctly in the FPI control software, hence the the necessity to check the centre regularly.

A prominent vignetting effect present in the background of a spectral profile could introduce bias to spectral lines centred at different wavelength-bins within the profile. If the background is much brighter for a fringe at a smaller radial distance from the centre of the image and drops away non-uniformly at a larger radial distance from the centre, a more inner fringe could be skewed leftwards more than a more outer fringe (or vice versa depending on the shape of the vignetting function). After fitting a Gaussian peak this can create an artificial red Doppler shift and peak width enhancement. Hence, this would

produce a spuriously enhanced positive vertical wind and temperature derivation for a more inner ring (or a more outer ring). The vignetting function for our FPIs needs to be investigated to understand how it disorts fringes centred at different radial distances away from the centre of the image.

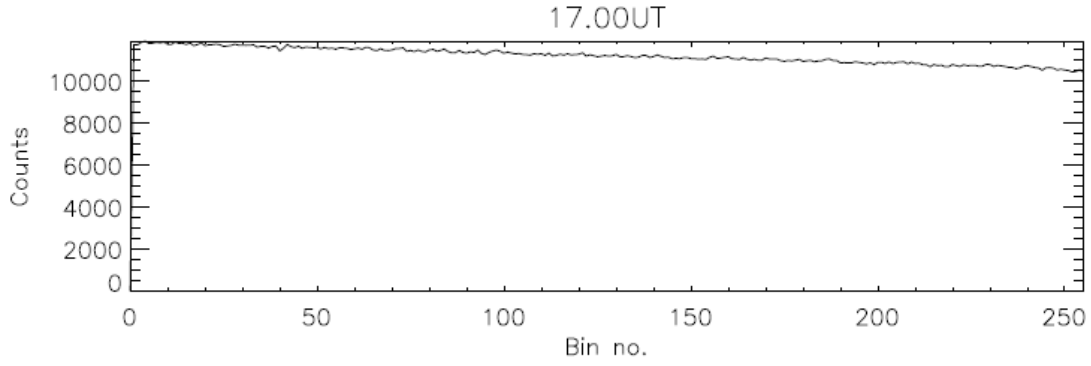


Figure 5.11: Response of the Kiruna red-line FPI to a flat field source of light; relative intensity counts vs. bin number

Injecting a purely flat field continuum, i.e., equal intensity light across the spectrum, into the FPI allows the vignetting effect to be observed in isolation with no superimposed spectral peaks. This flat field continuum response of the Kiruna FPI was investigated on 11th February 2015 on an instrument maintenance field trip. A flat field source, i.e., a standard incandescent torch bulb with equal intensity across the spectrum, was shone through the optics via the fibre optic channel used to inject laser light into the instument. The response of the instrument is shown in figure 5.11. Evidently, the resulting function is linear exhibiting the vignetting effect which is causing the intensity to decrease further away from the centre. The linear function shows that the vignetting will not affect the final wind calculations. This is because any systematic shift the vignetting effect will have on the sky spectra will be constant for all Doppler shifts, i.e, wherever the spectral peaks lie in the window, the same vignetting distortion will apply to each. This will manifest as a systematic shift in the calculated velocities. This would be removed from the data in the baseline removal procedure (see section 2.6.4).

## 5.5 OH Contamination

From the previous two subsections we have shown that the extreme vertical wind events present in Svalbard's midnight distribution's fat tail are unlikely to be real, and are also

unlikely to be instrumental effects. The other likely option to explore is the possibility of an additional contaminating line from another atmospheric species skewing the 630nm spectra.

Throughout our investigation so far we have come across many indications that an additional atmospheric species' emission line is contaminating the spectra and is responsible for the extreme vertical winds, e.g., the upwellings being coincident with temperature enhancements, although outside of the expected auroral heating region (figure 5.1); the asymmetric upturns in the temperature-intensity and vertical wind-intensity scatter plots (figures 5.5 and 5.7); and the off-zenith look directions exhibiting large divergences only when their intensities are low (figure 5.3). This section addresses spectral contamination directly, more specifically hydroxyl (OH) contamination. There is one other known species (besides OH) which emits close to the 630nm line and could potentially be contaminating the spectra: hot oxygen (4000K). It is found at ∼550km altitude and peaks at $60^o$ latitude, $300^o$ longitude at 24UT (Bessarab and Korenkov, 2011). Hot oxygen is found to not be as globally pervasive as hydroxyl and is only present at 1-2% the concentration of ambient oxygen. Hot oxygen would thus show up as a weak line in the spectra. Therefore, for this initial study, it is neglected. Hydroxyl, however, permeates the whole globe and is a well characterised atmospheric species. In this section we investigate this contaminant as an explanation for the spurious winds and temperature enhancements.

### 5.5.1 Atmospheric Hydroxyl

Hydroxyl (OH) is globally distributed across the mesosphere and most of the stratosphere. It is the most important oxidiser throughout the atmosphere (Conway et al, 1996) and is a key reacting species in the regulation of ozone (Minschwaner et al, 2011). There is a diurnal trend in OH concentrations which follows that of solar ultraviolet radiation (Wennberg et al, 2006) with the minimum at night-time falling to $1\times10^6\text{cm}^{-3}$ and the maximum around noon reaching $1\times10^7\text{cm}^{-3}$ (Conway et al, 1996). The seasonal trends in OH abundance are very minor, only varying by a few percent (Damiani et al, 2010). Since OH is more abundant at lower altitudes, measurements taken at larger zenith angles will view more of the species and will have a larger contamination. The key production reactions for OH are (Minschwaner et al, 2011):

$$H_2O + O(^1D) -> 2OH$$

$$H_2O + h\nu -> OH + H$$

and the key loss reaction is:

$$OH + HO_2 -> H_2O + O_2$$

where $HO_2$ is an atmospheric radical (Monks et al, 2005)

### 5.5.2 The Hydroxyl Lines in Spectra

The OH emissions have long been known as troublesome spectral contaminants for the FPI community's 630nm observers (Chamberlain et al, 1961; Hernandez et al, 1974; and earlier references therein). The greatly improved SNR of the spectra due to modern EMCCD detectors allows us to probe the OH signals in unprecedented detail. The three lines closest to the 630nm line with significant emission rates are at 628.7nm, 629.79nm and 630.7nm (vacuum wavelengths) which are for the respective OH transitions $P_1(2)$, $P_2(3)$ and $P_1(3)$. The Svalbard FPI filter has a 1nm bandwidth centred 0.2nm higher at 630.2nm to allow for off-axis rays to be detected (or fit through the bandwidth). Off-axis rays (maximum $\sim$ 7 degrees off-axis) view the central wavelength of the filter as 629.8nm. Allowing for both on-axis and off-axis rays, the filter only allows the second two lines, $P_2(3)$ and $P_1(3)$, to pass. This is demonstrated in figure 5.12.

Figure 5.12 simulates the response of the Svalbard FPI's spectrogram to the 630nm oxygen emission and all three nearby OH emission lines (only two fitting through the bandwidth) all representing species with zero Doppler shift. In the simulation, the filter is centred on bin 50 at a central wavelength of 630.2nm. The etalon gap is 18mm so the free-spectral-range (FSR) of the 630nm light is $1.1025\times10^{-11}$m from equation 2.9, and since a FSR covers 117 bins, this means a single bin covers a wavelength shift of $9.42\times10^{-14}$m ($\Delta\lambda_{bin}$), which represents $\sim$44ms$^{-1}$ (see equation 2.23). The maximum filter bandwidth including both on-axis and 7 degree off-axis rays ranges from 629.3-630.7nm, giving an overall bandwidth of 1.4nm ($\lambda_{Bandwidth}$). From the central wavelength of the the filter out to the edges of the bandwidth, the wavelengths are interpolated across each bin so that each bin is labelled with its corresponding wavelength. Since, the interference pattern is repeated every FSR (117 bins), the wavelengths at the end of the spectrogram need to be continued on the other end of the spectrogram. This is so that if there was an emission at a FSR more than 630nm, it would overlay exactly on the 630nm bin. Therefore,

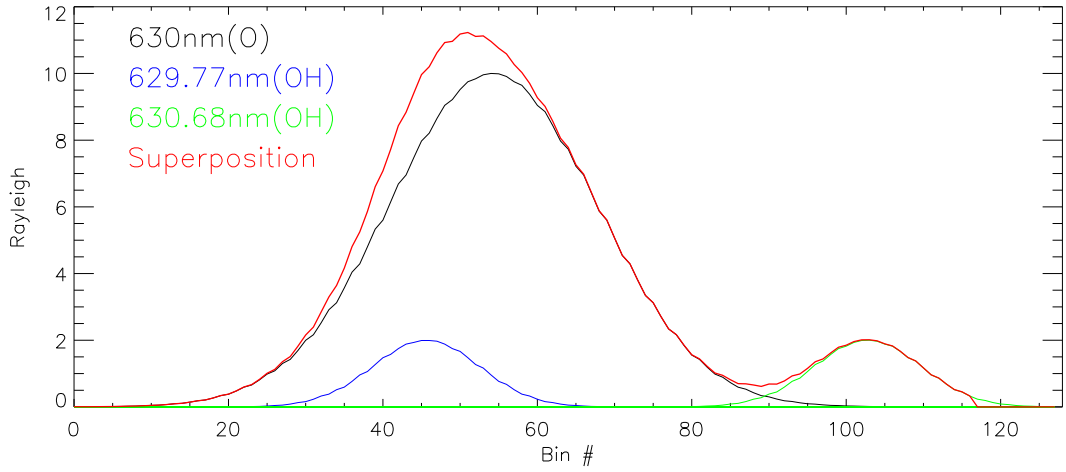


Figure 5.12: Example spectrogram simulation set-up for a FSR covering 117 bins and the 1nm wide filter centred on 630.2nm for on-axis rays and 629.8nm for 7 degree off-axis rays. The 630nm emission is 10R bright, centred at bin 54 and the two OH emissions fitting through the bandwidth are the 629.77nm and 630.68nm (air wavelengths) centred at bin 46 in blue and bin 103 in green. The red line over-plotted is the superposition of all three lines (to which the spectral fitting algorithm fits a Gaussian) centred at bin 51, i.e., 3 bins away from the pure 630nm emission. All lines plotted represent their species with a zero Doppler shift.

the wavelengths are interpolated (or 'wrapped') around the spectrogram every 117 bins ($\#bins(FSR)$) so that each bin corresponds to multiple possible wavelengths (across the bandwidth). The number of cycles of the FSR in order to cover the entire bandwidth, $\#cycles(FSR)$, is thus calculated by:

$$\#bins(bandwidth) = \frac{\lambda_{Bandwidth}}{\Delta\lambda_{bin}} =\sim 14862$$

$$\#cycles(FSR) = \frac{\#bins(bandwidth)}{\#bins(FSR)} = 127$$

Therefore figure 5.12 shows a spectral window composed of 127 overlaid plots of different wavelength ranges (on the x axis as spectral bins). The wavelength bins are assigned different intensity values subject to the different artificial emission lines introduced (see following), which make up the spectral peaks in the simulation of figure 5.12.

The different emission lines are introduced to the simulation by setting up Gaussians centred on the central wavelength plus any wavelength shift due to the Doppler shift (here zero). The width of the Gaussian is determined by the temperature of the emitting species (see equation 2.30) taken as an ambient 700K for the thermospheric oxygen and an ambient 200K for the mesospheric OH. The height of the spectral peak is taken to be the emission rate, i.e., the measure of emission brightness or the amplitude of the spectral line. In this case the 630nm emission is chosen to be 10R bright and the OH emissions

are 2R in brightness. The 630nm emission is centred at bin 54 and the two OH emissions fitting through the bandwidth are the 629.77nm and 630.68nm (air wavelengths) centred at bin 46 in blue and bin 103 in green. The red line over-plotted is the superposition of all three lines (to which a Gaussian function is fitted in the spectral fitting algorithm) with the main component centred at bin 51. The main peak of the net superposition line lies 3 bins away from the pure 630nm emission; which represents a change in wind velocity of $\sim$132ms$^{-1}$. The superposition peak is also sizeably broader than the pure 630nm peak which translates into an apparent temperature enhancement.

This simulation has demonstrated the effect that the OH lines have on displacing and broadening the measured spectra from which we derive wind and temperature values. Generally the OH intensities are much weaker than 630nm airglow and aurora so we are expecting it to only affect our parameter estimations a small fraction of the time. We need to simulate the 630nm line for various levels of intensity (which are relevant to the polar thermosphere) in order to vary the relative intensity between the OH and 630nm line. This will show us how low the 630nm intensities have to be in order for the OH line to affect the 630nm Gaussian fit significantly. In this work we will be using emission rates as a measure of brightness instead of intensity as this offers a simpler method of modelling the spectra (generally, emission rate$^2$ $\approx$intensity). Therefore the next step is to calibrate the 630nm emission rates so that they are absolute estimates.

### 5.5.3 Calibrating FPI Emission Rates with MSP

The Svalbard FPI dataset measures relative brightness values. We need absolute 630nm emission rates so that we may compare them with the characteristic emission rate of OH (2R) and attempt to quantify the level of contamination. The FPI emission rates are cross-calibrated with those of the collocated Meridian Scanning Photometer (MSP) which measure absolute 630nm zenith emission rate values in Rayleighs (see section 1.7.3). A relationship is found between the two data-sets in the form of some calibration function, most likely a straight line, which can be used to calibrate other data.

Figure 5.13 shows the relationship between the MSP brightness and FPI brightness for December 2007 data (the month of our extreme vertical wind case studies). There are two scatter plots of MSP emission rates against FPI emission rates; the top plot shows all the data and the bottom plot shows the lowest emission rate data. A strong correlation is found between the two entire datasets with an $R^2$ value of 0.87. This means that a

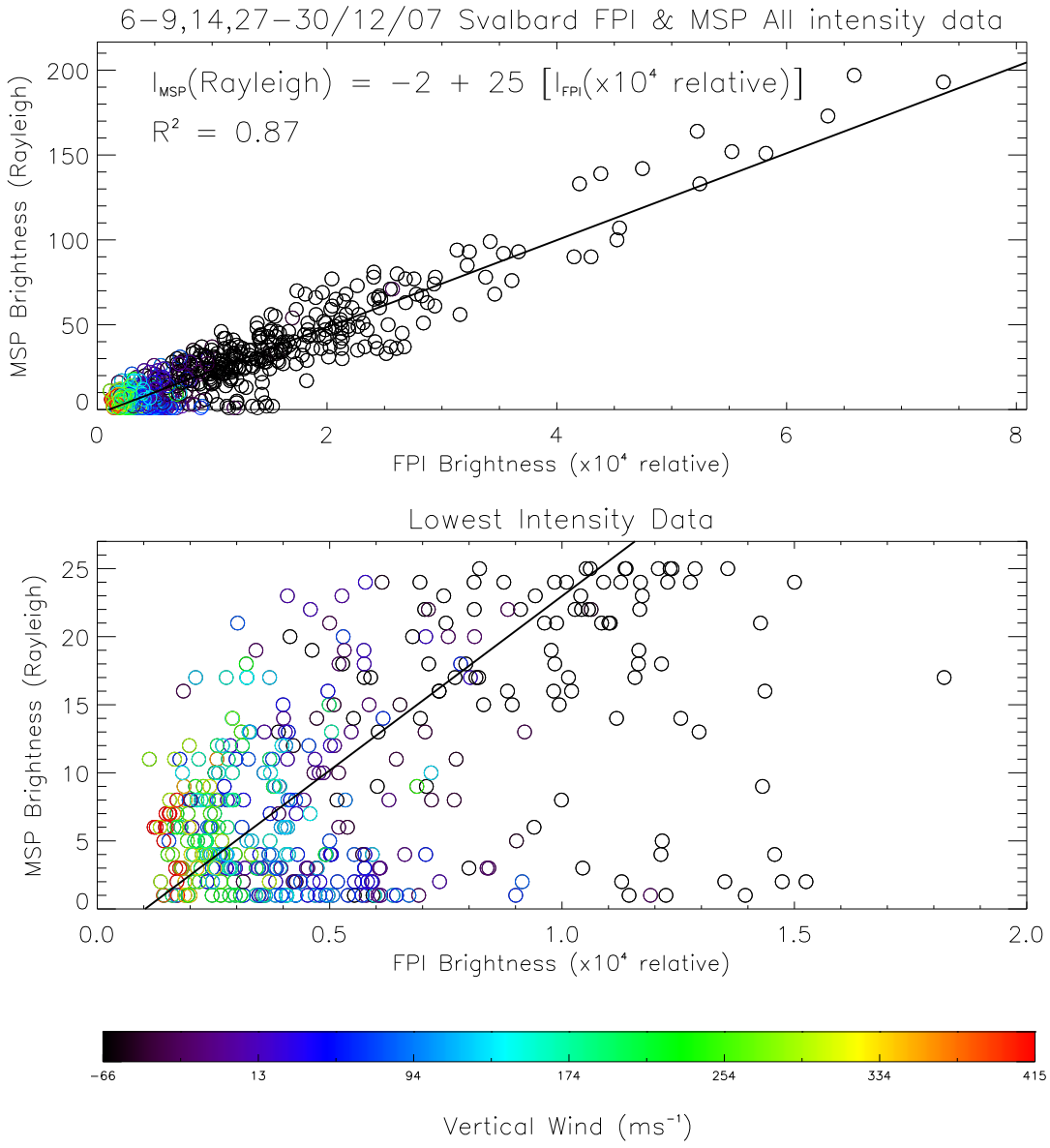


Figure 5.13: Scatter plots of MSP brightness (absolute Rayleighs) against FPI brightness ($\times 10^4$ relative units) colour coded by vertical wind velocity. The data used is from December 2007 (the month of our extreme vertical wind case studies). All the data are displayed in the top plot, only the lowest emission rate data are displayed in the bottom plot and the colour scale is shown in the colour bar underneath. A line is fitted to the data, the equation of which, is labelled on the top of the top plot as a calibration function for the FPI data. The $R^2$ values is labelled underneath showing a strong correlation between the two data-sets, validating the calibration function.

meaningful calibration function can be derived from the data as a straight line. A line is fitted to the data, the equation of which, is labelled on the top plot. To calibrate the FPI emission rates, we multipy by 25 and subtract 2. Due to the varying sensitivity of the FPI detector (see section 2.7) this calibration function will not apply to data from other seasons. In addition, the negative fitted intercept of the data implies that, a small proportion of the calibrated data will be calculated as small negative emission rates. These data are labelled zero in order to represent the lowest emission rates. When binning these values later, they will be included in the lowest value emission rate bin, i.e., 0-2R.

The data points are colour coded by vertical wind as shown by the colour bar underneath the plots. This shows the majority of the data to represent negative velocities; an artefact of an erroneous baseline correction on the data, due to the inclusion of the spurious upward winds in the calculation. Towards increasingly low emission rates, there is a gradient in colour matching that of the colour-bar. This represents a gradient in wind velocities to increasingly positive values. This can be seen in more detail in the bottom plot showing the maximum vertical wind $+400\text{ms}^{-1}$ values to occur below 10R. The vertical winds measured during periods of very low emission rates are the most heavily affected by the spectral contamination.

The calibration function is applied to all-clear-sky Svalbard FPI data from November 2007 to January 2008, which is a total of 45 clear sky nights. These data are displayed in the distribution in figure 5.14 with absolute emission rate as the x-axis and number of data points as the y-axis. The optimum bin-size is calculated to be 3R wide using the Freedman Diaconis rule. The data values range from <3-549R with a median of 57R. The majority of the data are distributed towards the smallest emission rates with an interquartile range of 22-118R (even thought maximum emission rate bin is 500R). The most frequent emission rate value is the second bin at 6R however this is only the 8th percentile of the data. We must bear in mind that these statistics are specific to the solar minimum winter of 2007, when we are seeing frequent extreme vertical wind events. Data from seasons farther away from solar minimum will exhibit an emission rate histogram with far more high emission rate values and far fewer low emission rate values. For example, on the CUSPN campaign in January 2012 (see chapter 7 figure 7.2) the absolute emission rates reach ∼10,000R.

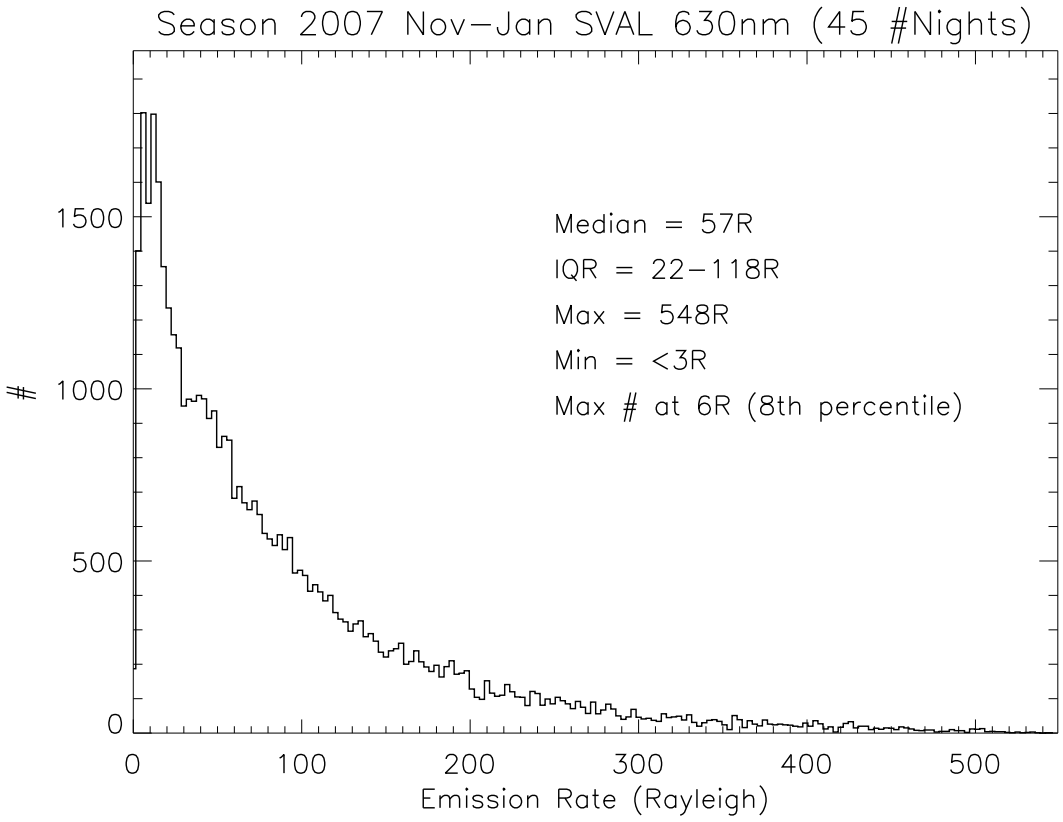


Figure 5.14: Histogram of MSP-calibrated Svalbard red-line FPI emission rates from season 2007 November-January clear sky data

### 5.5.4 Modelling the Impact of Hydroxyl Contamination

**Visual Impact**

For an introduction to the impact of OH contamination on our data analysis, we first consider it visually using artificial spectra. Simulations of the FPI spectrogram responses are carried out for the minimum, lower quartile, median and upper quartile of the 630nm emission rate values and are displayed in figure 5.15. These are in the same format as figure 5.12 except there is an extra line in purple showing the Gaussian fit to the superposition line.

The top plot shows a maximum 630nm emission rate value of 2R (the value of the lowest emission rate of our dataset), equal brightness to the OH lines. There is a clear distinction between all the peaks here. The Gaussian fit is significantly offset from the 630nm line by a few bins. When running these spectral peaks through the parameter analysis routine the OH here is adding a spurious $180\text{ms}^{-1}$ to the wind and 380K to the

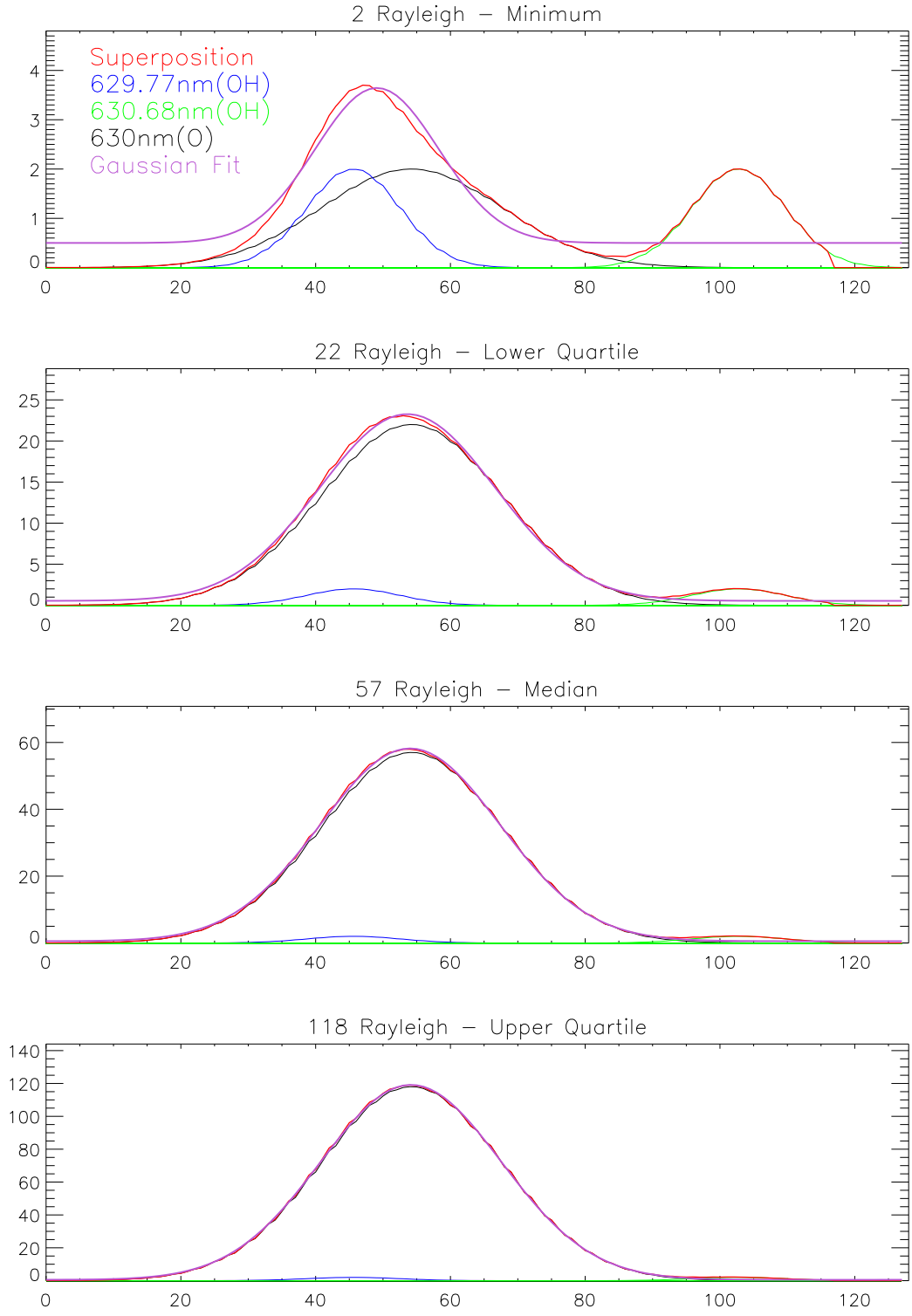


Figure 5.15: Simulations of the Svalbard FPI spectrograms for the minimum, lower quartile, median and upper quartile of 630nm emission rate data with two OH lines at 2R. This is formatted the same as figure 5.12 except with another line in purple representing the Gaussian fit to the superposition of lines.

temperature.

In the second plot, showing a 22R 630nm brightness (the value of the lower quartile emission rate of the dataset), the OH lines are small in comparison to the 630nm peak and the offset of the Gaussian fit is only <1 bin. The analysis routine shows the OH here adding a spurious 28ms$^{-1}$ to the wind and 67K to the temperature. Representing the lower quartile, this shows 25% of the dataset is affected to this extent or worse.

The third plot and fourth plot show 57R and 118R 630nm brightness, representing the median and upper quartile of the data. The OH lines and offsets of the Gaussian fit (from the pure 630nm peak) are hardly visible in both. The analysis routine shows the median (upper quartile) plot spectra affected by the OH adding a spurious 9ms$^{-1}$ (5ms$^{-1}$) to the wind and 22K (15K) to the temperature. The top 25% (50%) of emission rates will be contaminated to this degree or less (at least for the solar minimum 2007 winter data; other winters will have larger emission rates). These spurious vertical wind values are not insignificant as they are equivalent to their typical error bars (∼5-10ms$^{-1}$), whereas, the temperature contributions lie within their typical error bars (∼90K). The significance of these contributions is further discussed in the next two sections.

These plots illustrate the spectrogram's very high resolution; a small bin shift can thus result in a large wind velocity shift (44ms$^{-1}$ per bin). These differences need to be further quantified, in order to make analysis more accurate.

#### Typical Vertical Wind and Temperature Values

In the next section we quantify the impact of OH for a range of 630nm emission rates. However, before changes in vertical wind velocity and temperature due to OH can be analysed meaningfully we need to be aware of the typical magnitudes and standard deviations of these parameters within the polar thermosphere. Only by knowing the standard deviation expected of the data can we infer whether extra errors incurred (due to OH contamination) are significant or not in comparison with typical past measurements.

In section 5.3 the typical Svalbard vertical wind is inferred to be 0ms$^{-1}$, large winds to be 50ms$^{-1}$ and the standard deviation to be 15ms$^{-1}$. The average thermospheric temperature is known to be approximately 700K which has been demonstrated in the middle plot of figure 5.5 as a scatter plot of neutral temperature and SNR of signal. Figure 5.16 shows the distribution of neutral temperature values derived from 2006-2009 Svalbard FPI data. The data are positively skewed due to the spectral contamination

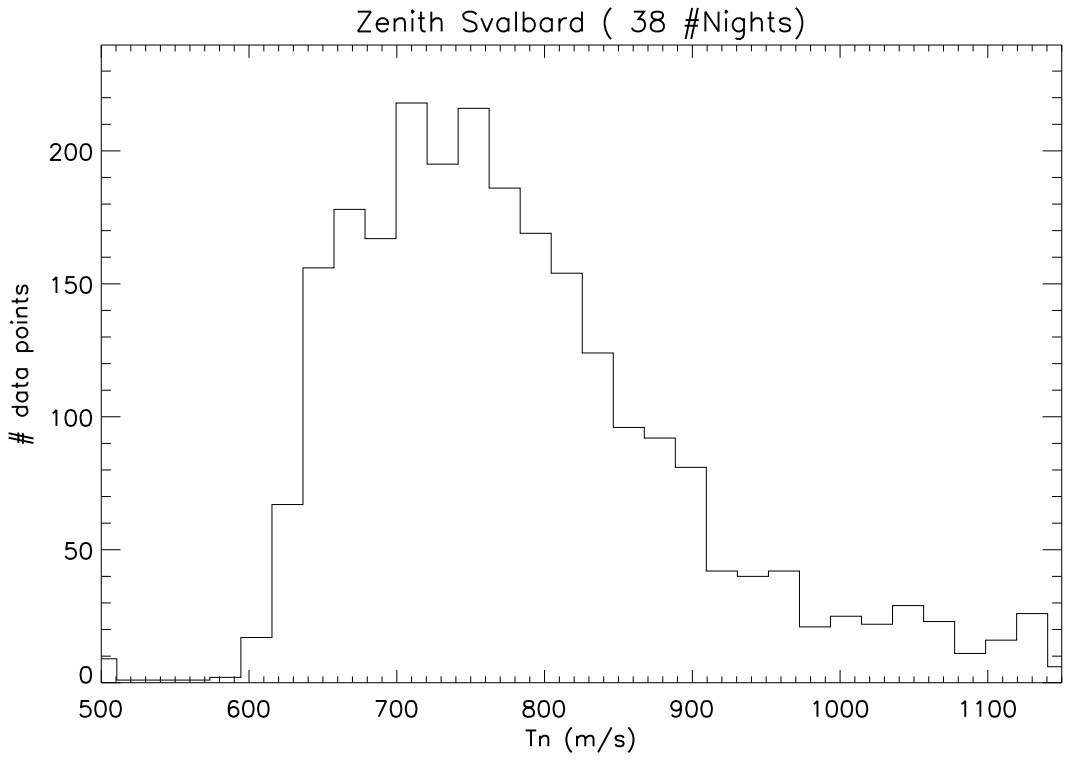


Figure 5.16: Neutral temperature frequency distribution from 2006-2009 November to January data derived from the 630nm Svalbard FPI measurements with a standard deviation of 115K.

which we have seen affects temperatures as well as wind velocity (see figure 5.1). The distribution peak is not well defined, however, it must lie between 710-752K (i.e., two peaks of distribution). The Svalbard FPI temperature observations' standard deviation is calculated to be ∼100K. This calculation has included data from the positive tail, which distorts the temperature population.

**Quantified Impact**

Simulations of the FPI spectrograms, similar to those illustrated in figure 5.15, are carried out for a range of 630nm emission rates between 0-500R. The FPI spectral fitting algorithm is performed on these spectra, i.e., the superposition of the 630nm and OH lines, and the wind and temperature values derived. This is repeated for all simulations with the OH lines removed. The differences in wind and temperature values are calculated between the spectrogram with and without the OH. The results are shown in figure 5.17. There are four plots showing the change in derived wind or temperature due to OH spectral contamination against 630nm brightness. The first two plots show the ranges 10-500R and the second

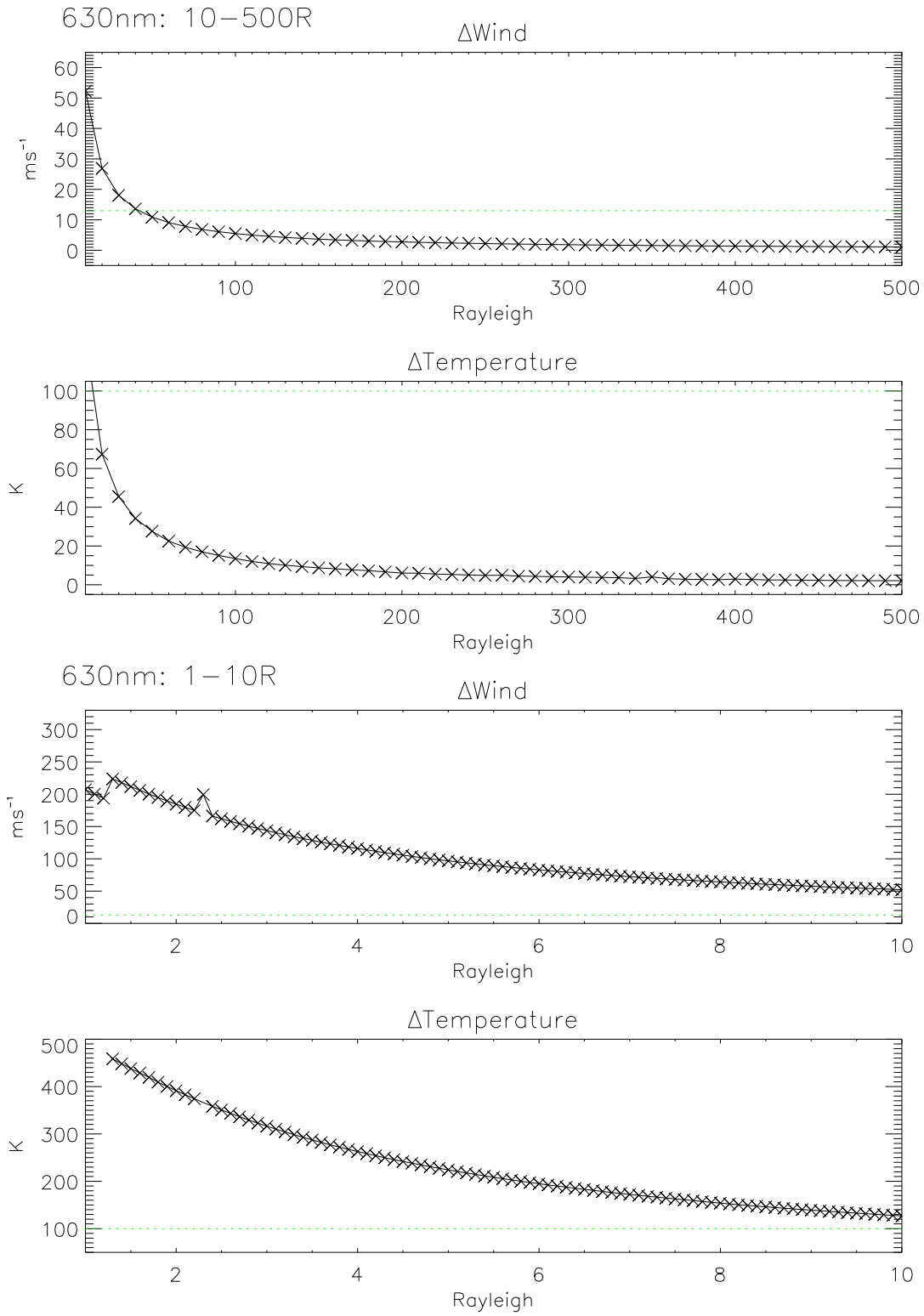


Figure 5.17: Four plots showing the change in derived wind or temperature due to OH spectral contamination for a range of different 630nm brightness values. The first two plots show the ranges 10-500R and the second two plots show the ranges 0-10R. Of each pair of plots, the top ones display the difference in wind velocity and the bottom ones display the difference in temperature. Values are derived from simulations of the Svalbard FPI spectrogram as displayed in figure 5.15.

two plots show the ranges 1-10R. Of each pair of plots, the top ones display the difference in wind velocity and the bottom ones display the difference in temperature.

Considering the 10-500R ranges (top two plots) both difference trends are fairly constant for most emission rates and then as zero brightness is approached the trends both dramatically increase. Both sets of data plateau at above ∼100R where the change in wind becomes smaller than 6ms$^{-1}$ and the change in temperature becomes smaller than 10K. For the upper quartile data, at 118R, the changes reduce to 4ms$^{-1}$ and 8K. At the median 630nm emission rate value (57R) the change in wind velocity is 11ms$^{-1}$ and the change in temperature is 16K. Moving to the lower quartile 630nm emission rate (22R) the change in wind increases to 30ms$^{-1}$ and the change in temperature to 45K.

Now considering the 1-10R data (bottom two plots) the trends show smooth curves tending to maximum changes, at 1R, of 320ms$^{-1}$ in wind and 500K in temperature, from minimum changes, at 10R, of 60ms$^{-1}$ in wind and 100K in temperature. The most frequent emission rate value of 6R shows changes of 100ms$^{-1}$ in wind velocity and 160K in temperature. This entire 1-10R range of data is thus heavily contaminated.

The offset in wind equal to the standard deviation of the vertical wind (15ms$^{-1}$) is reached when the 630nm emission rate drops below 50R. Whereas for the temperature, the change equal to its standard deviation (100K) is reached when the 630nm emission rate drops below 10R. Clearly, the temperature parameter is far more 'resistant' to changes due to the OH line than the wind velocity.

We define unusable data using the convention of any points outside the 1-sigma boundary. By calculating the percentiles from figure 5.14, 11% of the temperature data and 47% of the wind velocity data are unusable. These data are not discarded but are targeted for correction (see section 5.5.6).

### 5.5.5 Absolute 630nm Emission Rates at Midnight

Figure 5.18 shows the average absolute 630nm emission rates with UT measured by the Svalbard FPI from the 2007 season November to January. The error bars represent standard deviation; some dip below zero due to the calibration function (in figure 5.13) having a negative y intercept so that the lowest emission rate data points are calculated as negative. The negative values lie within the error of the intercept term and so it is reasonable to assume they are physical positive emission rate values. There is a minimum from 18-24UT about magnetic midnight reaching average values of 40R. Towards midday the trend in

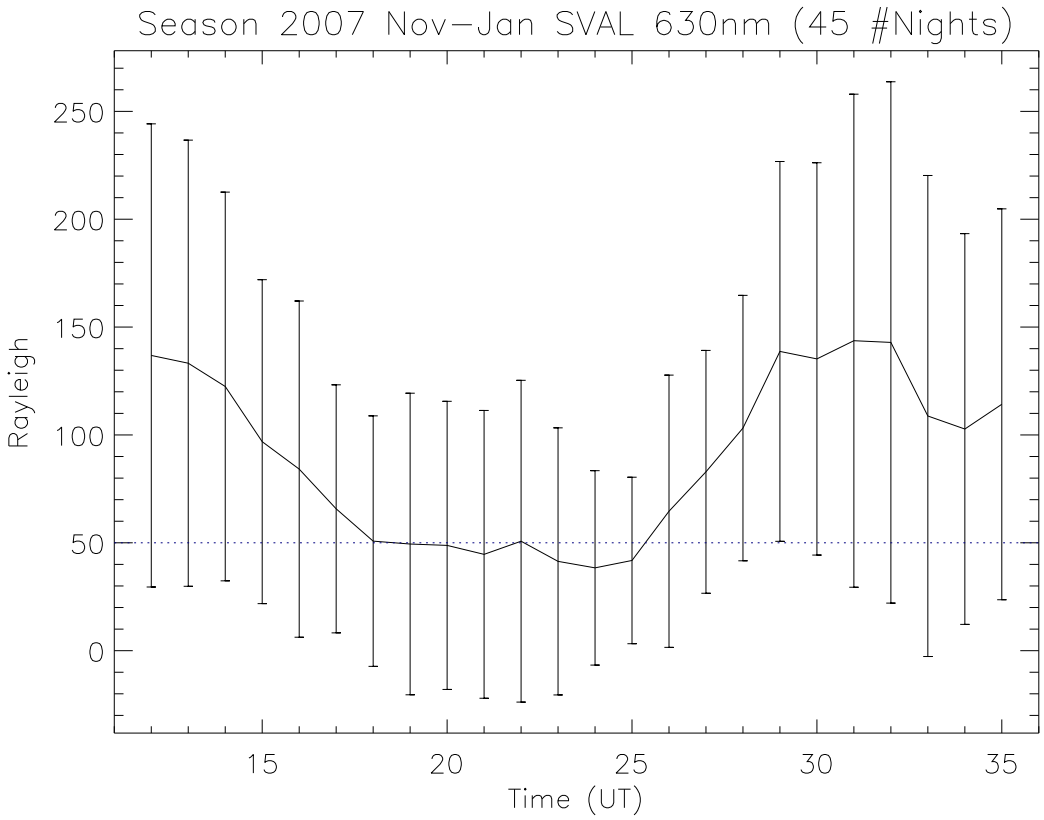


Figure 5.18: Average absolute 630nm emission rates with UT measured by the Svalbard FPI from the 2007 season November to January with error bars representing standard deviation within 1-hour UT bins.

average values increases to two peaks; one at 12UT of 140R and one at 32UT (11MLT) of 145R. The 'error' bars are large, indicating the large variability in the data. The blue dotted line marks 50R, the chosen threshold for winds being classified 'unusable'. From 18-25UT the emission rates are less than or equal to this threshold and so these are the hours on which we must be focussing our corrections. Having said this, the error bars are large and around this time, there is a good proportion of data above this value as well as below. Likewise, for all seemingly 'usable' data before and after this period, the errors show that a small proportion of these data can also reach emission rates below the 'usable' threshold.

Figure 5.19 displays four out of the six large vertical wind case studies which can be matched with MSP data; 14th, 28th, 29th and 30th December 2007. These plots are formatted identically to figure 5.1; the only difference between the two is the use of the absolute MSP emission rates instead of the relative FPI emission rates. The maximum

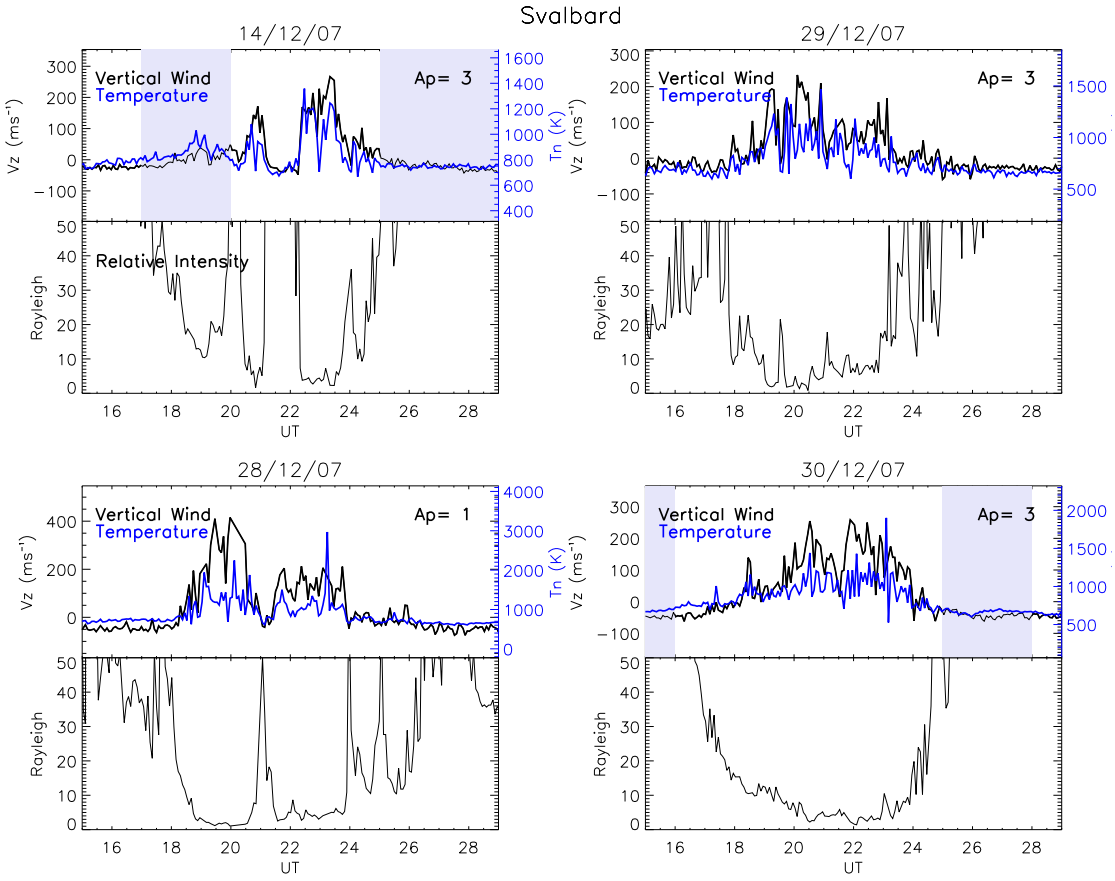


Figure 5.19: Line plots of four of the six extreme vertical wind case studies formatted as figure 5.1 with the emission rates as absolute MSP measured ones ranging from 0-50R

emission rate displayed on each graph is 50R.

Firstly, considering the vertical winds, whenever the emission rates dip below 50R, there is a perturbation in the wind magnitude. This perturbation dramatically increases the lower the emission rate. At the peaks of the upwellings, the emission rates are found to be below ∼5R when the change in vertical wind velocity becomes the order of 100ms$^{-1}$, as predicted in figure 5.17. Secondly, considering the temperature, there seem to only be significant enhancements of temperature at the times where the emission rates drop below 10R. These findings are all consistent with the relationship predicted by our simulations in figure 5.17. This is strong supporting evidence that the phenomenon we are observing is OH spectral contamination.

### 5.5.6 Correcting the Spectra

Measures need to be made in order to correct the data which involves accurate modelling of the OH emission on the spectrogram. The spectrogram simulations we ran are useful

to roughly gauge the magnitudes of spurious data (which match up with the magnitudes in Svalbard's fat positive tail) thus diagnosing Svalbard's fat positive tail as due to OH contamination. These simulations were based on all emission species having zero Doppler shift and the neutral temperatures (and hence spectral line widths) being the constant average values for the thermosphere (700K) and mesosphere (200K). A more precise analysis is required to remove the contamination from the spectra. In order to do this, a proper characterisation of the OH line is required, i.e., emission emission rate, neutral temperature and wind speed. This would involve knowing the state of the mesosphere to a very high spatial and temporal resolution which is infeasible due to measurement limitations. Therefore, assumptions must be made using climatologies of the wind and temperature in the mesosphere. This may not necessarily introduce much error. The reasons for this are now discussed.

Emission rates of the OH emission are well known to be weak with typical values of ∼2R (Meriwether, private communication). However the emission rates can vary between different transitions. An emission rate for the $P_1(3)$ line has been derived as 5-8R and for the $P_2(3)$ line to be 2-4R (Hernandez, 1974). Since these emission rate values are larger than the ones used in the simulations, using these emission rate values is likely to increase the threshold temperature and wind values of uncontaminated data.

The mesopause is the region where the atmosphere reaches its minimum temperature and also the winds are known to be particularly slow (Mullermann and Lubken, 2004). The OH layer lies at ∼87km altitude (Dyrland, 2013), just underneath the mesopause. Portnyagin et al (2009) use an empirical model to show that at about 90km altitude in November at $90^o$ latitude the vertical wind is of the order 0-2ms$^{-1}$. This would translate to a negligible shift in peak bin number for the OH line, therefore it is safe to neglect the Doppler shift of the OH line (as we have done in the spectrogram simulations).

The temperature of the winter mesosphere at the peak altitude of OH is typically ∼200K (Damiani et al, 2010); however, this layer has also been shown to reach 220K (Epsy and Stegman, 2002). These temperature values would translate into a peak width much smaller in comparison to the 630nm width, as we have seen. In our simulation, we chose a typical 200K for the mesospheric temperature value. Choosing a larger value, e.g., 220K, will contaminate the data more strongly and thus increase the threshold 630nm emission rate for uncontaminated data. The impact that a change in mesospheric temperature or emission rate could have on the derived parameter values needs to be investigated further.

### 5.5.7 The Cleaned Vertical Wind Histogram

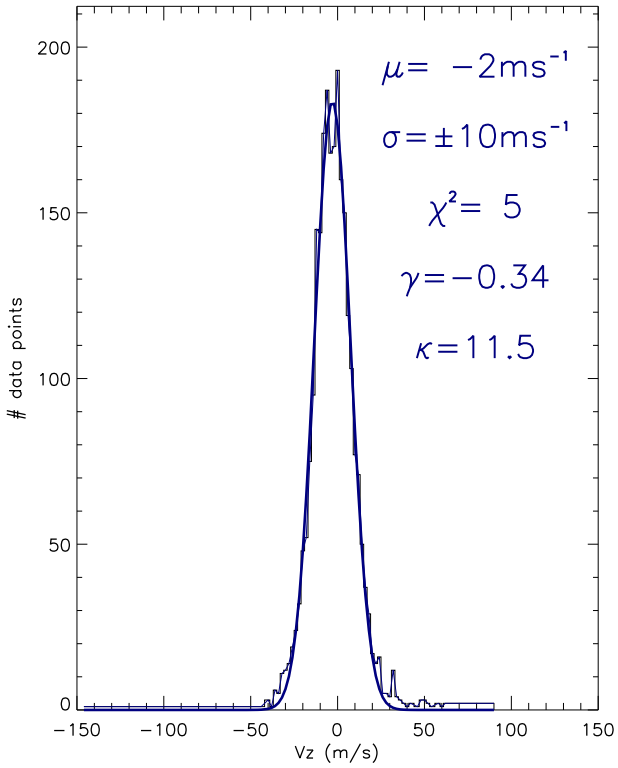


Figure 5.20: Svalbard's clean vertical wind distribution from 2007 with the OH contaminated data removed before the baseline is calculated.

The Svalbard vertical wind raw data have been recalculated excluding all unusable data so that they do not affect the baseline. The data are shown as a distribution in figure 5.20. The distribution is now more symmetric and centred more about zero. The fat positive tail has been drastically reduced although there are still a small amount of extreme vertical winds occurring between 50-100ms$^{-1}$. This could be due to the modelled threshold for OH contamination being too low. This would likely be improved with more refined spectrogram modelling. These are likely to be the cause of the distribution being centred on -2ms$^{-1}$ instead of zero due to a spurious baseline calculation. The standard deviation is reduced from an estimated 15ms$^{-1}$, in figure 5.3, to 10ms$^{-1}$. These are typical values of vertical wind observed at mid and low latitudes (Meriwether, 2011).

Removing contaminated data using the modelled OH-contamination emission rate threshold results in a more symmetric vertical wind distribution that is centred on zero. This is now consistent with the Kiruna distribution, and with the anticipated convective behaviour on the polar cap edge of the auroral oval. The improved distribution gives

confidence that the Svalbard fat positive tail is the result of OH contamination.

### 5.5.8 Conclusion

In conclusion it seems highly likely that OH contamination is the culprit for the Svalbard vertical wind distribution's fat positive tail. From our initial modelling of the contamination, it seems any 630nm data with an emission rate less than 50R (10R) has unusable wind (temperature) values and requires correction. This amounts to 47% of the vertical wind data and 11% of the temperatures from the 2007 season. This is expected to be the worst affected year due to the poor signal associated with solar minimum. A more sophisticated simulation of the OH line needs to be carried out in order to correct the entire database.

### 5.5.9 Future Work

- The entire FPI database needs to be cleaned of OH contamination (calibrating each year with MSP data). A new step in the raw FPI data analysis routine needs to be introduced involving calibrating the FPI emission rates with the MSP and finding the new threshold values for unusable data. This process needs to be carried out at the start of each observing season to track the drift in the FPI's relative emission rates.

- With the cleaned vertical wind data the average vertical wind MLT variation needs to be plotted to look for signatures of the cusp upwelling

- A study is required for the presence of the OH in the off-zenith look directions spectra which we also found to be contaminated. This contamination is likely to be worse as these viewing volumes are at low elevation angles where a greater proportion of the lower atmosphere is sampled (see section 3.9.1) and the OH line is likely to be stronger.

- A more sophisticated correction algorithm needs to be written to identify the OH line on each spectral profile.

- The algorithm needs to be tested by comparing corrected data with the collocated EISCAT Svalbard Radar (ESR) dataset at geomagnetically quiet times when the winds and temperatures are of similar magnitudes.

- Hot oxygen could be simulated to see if this has an impact on the derived data. It is unlikely to have as large an impact on the FPI data as OH does due to its relatively low concentration. It is likely to affect the temperature data more than the wind data as it will have a wide spectral profile indicative of the 'hot' oxygen (>4000K). Its expected low-amplitude peak is unlikely to noticeably skew the 630nm spectra.
- This whole study could be repeated for the KEOPS database to rule out spectral contamination, possibly contributing to the leptokurticity of the distributions.

# Chapter 6

# Burnside relation

The Burnside relation is a very simplified model of the thermospheric wind relating the vertical to horizontal components as:

$$\vec{w} = H(\vec{\nabla} \cdot \vec{u}) \tag{6.1}$$

where $\vec{w}$ is the vertical wind, $H$ is the pressure scale height and $\vec{u}$ is the horizontal component of the neutral wind vector. This relationship exists due to conservation of mass, i.e., with a hydrostatic atmosphere, the mass in a column of gas above a given pressure level will be constant. An upward wind would cause the mass to increase in the column above meaning that there must be a divergence in the horizontal wind somewhere along the column to compensate; and vice versa, with a mass-depleting downward wind there must be a mass-conserving convergence somewhere above in the column.

Over the years, this relationship has been of great interest to the FPI community as a tool for deriving the vertical wind values indirectly from the horizontal wind measurements. Historically, before the use of CCDs (pre-2000s), this capability was particularly useful since the FPI detectors were not sensitive enough to differentiate the vertical wind's small magnitudes (typically within $\pm 10 \text{ms}^{-1}$) from the measurement error. The horizontal winds, however, have much larger magnitudes ($\sim 100 \text{ms}^{-1}$), so they have always been easy to measure more accurately.

Nowadays, it is still useful for the FPI community to be able to determine vertical winds indirectly as it is not possible to measure the vertical winds in off-zenith look

directions (30-45$^o$ elevation). The vertical component in the off-zenith look directions is too small, whereas the horizontal winds have a large component so can be estimated to a good precision. A knowledge of the vertical wind in the off-zenith look directions is important to gain an understanding of the horizontal scale sizes in the polar cap. In particular, for our Svalbard FPIs, this knowledge would be a useful asset to the SCANDI horizontal wind-fitting procedure where assumptions are made about the horizontal scale of the vertical wind (see section 3.11).

For the general space weather community, vertical winds are relevant as they are key in the dynamics of the thermosphere, transporting species out of chemical equilibrium, consequently affecting the altitude of the ionosphere and distributing heat and energy, and thereby affecting the density structure of the atmosphere. Hence, reliable wind measurements should be incorporated into satellite models to improve high-precision orbit determination, and also into atmospheric models, in order to understand situations when hydrostatic equilibrium may break down.

In this chapter we will review previous work on the Burnside relation, derive the relation, present a statistical study to test the relationship on our dataset and then go through the assumptions of the relation to understand its shortcomings.

## 6.1 Previous work

The original work on the Burnside relation was by Burnside et al (1981). In their work they derive the relationship specifically for Arecibo Observatory (18$^o$N, 67$^o$W), a low latitude observing site, where vertical winds are typically too small to measure directly, $\sim$5ms$^{-1}$, so this equation enables them to be calculated indirectly. They expect a ten fold increase in precision of vertical winds calculated using the Burnside relation rather than from direct measurement at the zenith. This is estimated by comparing the relative magnitudes of the vertical wind component and horizontal wind divergence component within a Fourier decomposition of the wind field.

The Burnside et al (1981) work motivated many investigations into this relationship at different latitudes. Biondi et al (1984) was the first, using a mid-latitude observing site to compare the vertical wind to the divergence. They find divergent flows (positive $\vec{\nabla} \cdot \vec{u}$) present with upwards winds and convergent flows (negative $\vec{\nabla} \cdot \vec{u}$) present with downward wind, however the magnitude of the vertical wind is much larger than the model predicts.

Smith and Hernandez (1995) examine this relationship in the polar cap upper and lower thermosphere over a 13-day geomagnetic storm period using South Pole FPI data. In the upper thermosphere they find a consistent negative linear relationship between vertical wind and divergence throughout the storm period, violating the Burnside relation. In the lower thermosphere they find, only on the most active night, a negative correlation, again violating the relation. This is attributed to the larger scales of motion under higher geomagnetic activity causing the lower thermosphere to behave more like the upper. Although negative, the magnitude of the derived scale heights are approximately double the expected values at these altitudes. The negative linear relationship is explained by the vertical winds being driven by the divergence rather than the other way round which was argued by Burnside et al (1981). It is unsurprising that Burnside-type behaviour was not observed here as the equation is limited to quiet-time, small-scale vertical winds. The remaining studies on this relation (Guo and McEwen, 2003; Anderson, 2011; Anderson, 2011 (thesis); Ishii et al, 1994; Crickmore, 1991; Cooper et al, 2009) have used data from geomagnetically quiet nights.

Guo and McEwen (2003) average five years of central polar cap FPI data and find a positive linear relationship between vertical wind and horizontal divergence, with stronger divergence variability associated with the larger vertical winds. A crude calculation of the gradient gives a scale height of approximately 200km, over four times what is expected at this altitude.

Anderson (thesis 2011) uses five years of auroral latitude southern hemisphere FPI data with two years of SDI data and reports, under quiet conditions, a positive linear relationship between the average vertical wind and the horizontal divergence. The trends looked largely linear for smaller magnitude divergence values, while, for the larger magnitude divergence values, the scatter was significantly larger. He finds linear fits displaying Burnside-type behaviour with apparent scale heights ranging from 28.3km to 56.2km, compared to an expected scale height of $\sim$37km. In another work, Anderson et al (2011) study two geomagnetically quiet nights for the same observing site and also find a positive linear relationship. The gradient for each night's data give apparent scale heights of 99.5km and 112.4km, 3-4 times greater than expected. There is also a consistent large scatter on the plots, more so for the larger magnitude vertical winds, just as in the previous Anderson (thesis 2011) work and Guo and McEwan (2003) work. This type of scatter indicates that the assumptions underlying the Burnside equation are often, and at least locally, violated

with large wind magnitudes.

Crickmore et al (1991) perform a statistical analysis on three winters of high-latitude southern-hemisphere observations (excluding any times where aurorae were visible) and consistently find a positive linear relationship between vertical wind and divergence with an average apparent scale height of 225km, approximately 5 times what is expected in the real atmosphere. They find that application of the Burnside method to derive the vertical wind from the divergence predicts a vertical wind of typically half that actually measured.

Ishii et al (1999) use observations from both an SDI and an FPI at Ramfjord, Norway, and find a vertical wind-divergence relationship for one day to be time dependent. They perform a cross-correlation between the two parameters at each MLT with a time lag of $\pm$30mins to find that the correlation coefficient is negative most of the time. They explain this finding with the argument that the divergences drive the vertical wind. No attempt is made to calculate a scale height.

Cooper et al (2009) test the Burnside equation by using a local-scale, time-dependent 3D model of the neutral atmosphere to find that, at high latitudes, under stable conditions, the Burnside relation is only applicable above the height of maximal energy deposition and during times when forcings do not change rapidly with time. Below this region Cooper et al (2009) find the relationship between the vertical wind and divergence to yield a negative value for the scale height. Due to the fact that the height of maximum energy deposition is usually the E region, this validates the use of 630nm airglow emission observations to test the Burnside relation, at least during periods when the energy deposition is not changing rapidly with time.

Smith and Hernandez's (1984) negative scale height and Ishii et al's (1999) negative correlations are the odd ones out and these too are the only ones using data from geomagnetically active periods. In both works it is suggested that the negative sign is due to the divergence (formed by some horizontal driving force e.g. enhanced ion drag) being the cause of the vertical wind (rather than the former theory, developed by Burnside et al (1981), that the vertical motion was the cause of the divergence). The negative correlations demonstrate non-Burnside mass flows, i.e., an upwelling concurrent with a divergence and a downwelling concurrent with a convergence. These violate the zero mass flow boundary condition at the top of the atmosphere within the Burnside equation derivation.

The picture that emerges is that a linear trend is always detected; negative trends usually emerge from active-time datasets and positive trends from quiet-time datasets.

Most works find a positive scale height, however, the magnitude is typically 3-5 times too large, apart from the work by Anderson (thesis 2011), showing scale heights of realistic magnitudes.

## 6.2 Burnside derivation

Burnside et al (1981) derived this relationship in order to measure indirectly the small, low-latitude vertical winds from horizontal wind measurements. Within the work, Burnside et al (1981) refer to Hess (1959) for the derivation of the tendency equation, a step in the derivation of the Burnside relation.

### 6.2.1 Deriving the Tendency equation

In order to drive the vertical winds, divergences must be present or vice versa, i.e., the mass continuity equation must be satisfied:

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0 \tag{6.2}$$

The tendency equation assumes hydrostatic equilibrium (in integral form with limits representing altitude, z, from the emission height, z=h, to the top of the atmosphere, z=$\infty$):

$$P = \int_h^\infty \rho g dz \tag{6.3}$$

where we have assumed constant gravity, which is valid over the domain of the atmosphere.

The time derivative of equation 6.3 is:

$$\frac{\partial P}{\partial t} = g \int_h^\infty \frac{\partial \rho}{\partial t} dz \tag{6.4}$$

Substituting the mass continuity equation into this we get the rate of change of pressure in terms of the integral of the divergence of the winds:

$$\frac{\partial P}{\partial t} = -g \int_h^\infty (\frac{\partial (\rho u)}{\partial x} + \frac{\partial (\rho v)}{\partial y} + \frac{\partial (\rho w)}{\partial z}) dz \tag{6.5}$$

where x and y are the zonal and meridional directions. The vertical term of the above equation can be integrated directly using the boundary condition that the mass flow at

$z = \infty$ is zero (since the density must be zero there). The variation of $\rho$ and $w$ in the horizontal plane is negligible compared to the vertical direction, therefore, $\rho w = \rho(z)w(z)$ and the intergral of the vertical term can be written:

$$-g \int_{z=h}^{z=\infty} \frac{d(\rho w)}{dz} dz = -g \int_{\rho w}^{\rho w=0} d(\rho w) = +g(\rho w)_h \tag{6.6}$$

Then the tendency equation is derived by substituting this vertical wind term into equation 6.5

$$\frac{\partial P}{\partial t} = -g \int_h^{\infty} (\frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y}) dz + g(\rho w)_h \tag{6.7}$$

The tendency equation states that the pressure at a given point depends only on the weight of the air in the column above it, therefore the rate of change of pressure is equal to the rate of change of weight within the column. Since the density of the mass flow at the top of the column (at $z = \infty$) is zero, there can be no vertical wind flowing through the top, only the bottom. This means the weight of the column can only change via two processes: mass flowing through the sides, i.e. integrated horizontal divergence term, or mass flowing through the bottom, i.e. the vertical wind term. This step obviously ignores the non-physicality of wind at infinity.

### 6.2.2 From the Tendency Equation to the Burnside Relation

From the tendency equation Burnside et al (1981) made two assumptions about the atmosphere to even further simplify the behaviour of the atmosphere. It was assumed that the horizontal velocity is constant with altitude and that the time-dependent rate of change of pressure at the emission height is negligible.

Expanding the horizontal derivative gives:

$$\frac{\partial P}{\partial t} = -g \int_h^{\infty} \rho(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) dz - g \int_h^{\infty} u \frac{\partial \rho}{\partial x} dz - g \int_h^{\infty} v \frac{\partial \rho}{\partial y} dz + g(\rho w)_h \tag{6.8}$$

Assuming that the horizontal velocity and its horizontal gradient is constant with altitude, the velocities can be taken outside of the integral over dz. Then rearranging in terms of $-g\rho$.

$$\frac{\partial P}{\partial t} = (\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) \int_h^{\infty} (-g\rho) dz + u \int_h^{\infty} \frac{\partial(-g\rho)}{\partial x} dz + v \int_h^{\infty} \frac{\partial(-g\rho)}{\partial y} dz + g(\rho w)_h \tag{6.9}$$

Substituting back in the equation for hydrostatic equilibrium and reversing the order of integration and differentiation gives:

$$\frac{\partial P}{\partial t} = (\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) \int_P^0 dP + u\frac{\partial}{\partial x}\int_P^0 dP + v\frac{\partial}{\partial y}\int_P^0 dP + g(\rho w)_h \tag{6.10}$$

Then by neglecting the horizontal derivatives of pressure, and performing the pressure integrals, gives the rate of change of pressure:

$$\frac{\partial P}{\partial t} = -P(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) + g(\rho w)_h \tag{6.11}$$

Using the assumption that the rate of change of pressure at the emission height is negligible gives:

$$P(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) = g(\rho w)_h \tag{6.12}$$

Finally dividing by $g\rho$ and using the relation that the scale height H is equal to $H = \frac{P}{g\rho}$ gives the Burnside relation below:

$$w = H(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) \tag{6.13}$$

Since $H = \frac{P}{g\rho} = \frac{kT}{Mg}$ and we assume that within a scale height all atmospheric parameters are constant, temperature

The assumptions made within this derivation are:

- hydrostatic equilibrium
- zero mass flow at the top of the atmosphere
- horizontal velocity (and horizontal gradients) constant with altitude
- rate of change of pressure with time at emission height is negligible

## 6.3 Analysis

The validity of the Burnside relation to model the polar cap thermosphere at Svalbard is tested with a statistical analysis of SCANDI's 630nm atomic oxygen airglow observations together with the Svalbard FPI's 630nm measurements. The FPI zenith zone line of sight

measurement is taken to be the vertical wind and the horizontal winds are derived by fitting a horizontal wind field to the SCANDI line of sight winds using Conde et al's (2001) method as described in section 3.11.

### 6.3.1 Vertical Wind Contamination of Line of Sight Wind

The horizontal winds are estimated using the off zenith line of sight measurements, $U_{los}$ which are composed of a vertical as well as a horizontal component, as follows:

$$U_{los} = U_h \cos(30^\circ) + U_z sin(30^\circ) \tag{6.14}$$

where $U_h$ and $U_z$ are the horizontal and vertical components of the wind and 30° corresponds to the zenith angle of Svalbard FPI's off-zenith look directions. This entanglement of the two components in $U_{los}$ complicates the study, which requires the vertical and horizontal wind fields to be distinct so that they may be compared. It follows that if the vertical wind component is not correctly removed from the line of sight wind then it will contaminate the calculated divergences and could lead to an incorrect relationship between the vertical wind and divergence.

As shown in Crickmore (1993), supposing a 30° elevation angle and an all-sky vertical wind $U_z$, the vertical wind will contribute a $U_z sin(30^\circ)$ component to the line of sight wind measurement. This would result in the horizontal wind being offset from the actual value by $\frac{U_z sin(30)}{sin(60)}$ =0.577 $U_z ms^{-1}$.

Consider the same vertical wind component at two different observing regions in the sky, both at the same elevation angle but separated azimuthally by 180 degrees. In other words, zones on opposite sides of the FOV. The vertical wind will add 0.577 $U_z ms^{-1}$ to each region's radial (positive outwards) horizontal wind measurement. If the divergence is crudely calculated by subtracting the difference in the horizontal winds from these regions and dividing by the distance between them, then it is apparent that two times this vertical wind component contaminates the divergence.

Now calculating the divergence as described, assuming an observing region 433km away from the zenith viewing volume, gives:

$$\frac{2 \times (0.577 U_z ms^{-1})}{433km} = 2.67 \times 10^{-6} U_z s^{-1} \tag{6.15}$$

For an extreme vertical wind event of 70ms$^{-1}$ this gives a spurious 1.87 $\times 10^{-4}$ s$^{-1}$ addition to the divergence. This represents a significant source of error. The above equation shows that the larger the positive vertical wind, the more divergence will 'appear' and the larger the magnitude of negative winds the more convergence will 'appear'. This effect will manifest itself in the correlation of the vertical wind/divergence as a positive relationship, consistent with the Burnside relation. Therefore it is essential that the vertical wind component is accurately removed from the line-of-sight measurements for a fair testing of the Burnside relation to proceed.

### 6.3.2 Vertical Wind Assumption

The two ways to treat the vertical wind component present in off-zenith line of sight wind measurements are:

- neglect it, as the vertical winds are typically so much smaller in magnitude than the horizontal (magntiudes of -2$\pm$10ms$^{-1}$ for Svalbard (figure 5.20))
- assume that it is constant across the whole field of view and equal to the zenith measurement and then subtract it from all off-zenith measurements

The latter approach is adopted, with the zenith zone wind's LOS component subtracted from all zones. This allows the inner rings of SCANDI, which are close to the zenith vertical wind measurement, to be correctly adjusted, and possibly adequate adjustments for the outer ring measurements.

For the inner rings, the vertical component of the LOS is dominant and, assuming these rings share the same vertical wind as the zenith zone, the vertical component can be successfully removed. This leaves the minor but accurate estimate of the horizontal component. For the outer rings, this correction is not likely to be so successful. It is likely the horizontal vertical wind scale size is smaller than SCANDI's FOV so the vertical wind is likely to be different to the zenith value $\sim$500km away. Therefore, correcting the outer ring by removing the zenith wind LOS component is likely to introduce an error. However, this error is likely to be small compared with the large horizontal component present in large-zenith-angle measurements.

By averaging over the entire FOV, the merits and drawbacks of correcting for the inner and outer rings are combined to give a best estimate of the bulk flow of wind. Furthermore,

a median filter is applied to all wind signals before divergences are calculated to ensure we are looking at bulk motion and not localised fluctuations.

### 6.3.3 Calculating Divergence

A manual calculation of the divergence from zone to zone is performed on the final best-estimate of the horizontal wind field. As this wind field is a combination of the fitted perpendicular to LOS component and the measured LOS component, there will be real measurements included into the divergence as well as having separate values for each zone allowing more fine scale structure to be detected.

The method involves:

- interpolating the wind field radially and azimuthally
- calculating the radial and azimuthal changes in meridional ($dv$) and zonal ($du$) velocity from zone to zone
- calculating the distance between the zones radially and azimuthally ($d\theta$ and $dr$)
- calculating divergence in polar coordinates
- averaging divergences over the FOV to give the optimum all-sky divergence value

### 6.3.4 Statistical Analysis Method

All clear-sky vertical wind data (from the FPI) is paired with the corresponding divergence data (from SCANDI) from 2008-2014. The data known to be contaminated by OH (see chapter 5), is excluded from the dataset. The divergence data are interpolated to the lower time resolution of the FPI. A 2-hour wide sliding window is used to select data intervals to test. For each interval a linear model is fitted between the divergence and vertical wind which, according to the Burnside relation, yields the pressure scale height as the gradient. From the fit, three quantities are taken for this statistical study:

1. the coefficient of determination (R) to determine whether the linear trend is positive or negative, i.e., complying with or violating the Burnside relation
2. the correlation coefficient ($R^2$) as the measure of goodness of fit to decide whether the estimate of scale height is reliable enough to include in the statistical analysis

3. the apparent scale height from the gradient to see if it complies with measurements (∼50km) and, if not, to give an idea of the new relationship SCANDI is detecting

The three above quantities are displayed in histograms, seen in figure 6.1 and discussed in the following section.

## 6.4 Results

Figure 6.1 presents all the relevant correlations calculated for the SCANDI dataset. The top histogram shows the frequency distribution of the correlation coefficients. It shows the majority of all correlations (83%) were positive which complies with the Burnside relation. This leaves 17% of all correlations violating the Burnside relation's upper boundary condition, i.e., a convergence with an upward wind will incur mass flow at the top of the atmosphere, or conversely, a divergence and a downward wind would need to be fed by a mass flow at the top of the atmosphere. There is no data in the top bin due to there being no correlations strong enough to have R>0.95.

The second histogram then displays the frequency distribution of the coefficient of determination (or $R^2$ value) for a reduced dataset. Here any negative correlations are excluded which amounts to 17% of the data-set. The different intervals of the coefficient of determination, marked with dot-dash lines, represent how well the correlations are fitted by the linear model. They are split by rules of thumb into no relationship found ($R^2 < 0.3$), mild ($0.3< R^2 <0.5$), moderate ($0.5< R^2 <0.7$) and strong ($0.7< R^2 <1$) relationships. No relationship is found in 31% of the dataset, a mild relationship was found in 12%, a moderate relationship is found in 15%, leaving 25% of the data showing a strong relationship.

The final histogram displays the frequency distribution of the apparent scale height from the relationship between the vertical wind and divergence i.e. the gradient of the linear model fitted to the data (according to Burnside 1981). The only data included from the previous histogram are the ones which show a moderate relationship or better, leaving 61% to be excluded from the original dataset. The histogram displays a near-Gaussian distribution. There is quite a large fraction of the data with the FWHM of roughly 300km. Nearly all of the values are far larger than is expected in the real atmosphere, where the scale height is often assumed to be approximately 50km.

The MSIS model (see section 1.8.2 in chapter 1) was used to find the range of scale

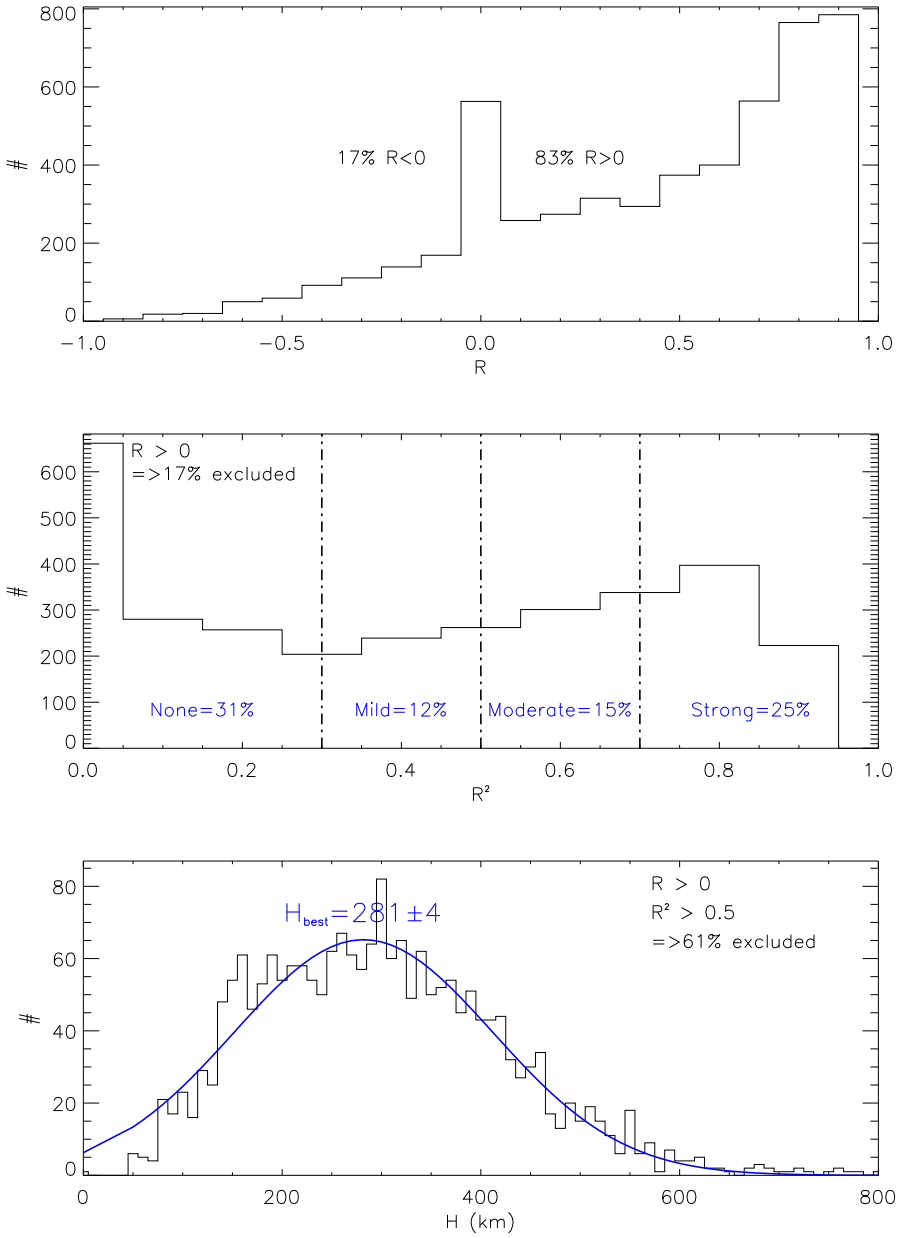


Figure 6.1: Histograms showing the relationship between Svalbard FPI vertical wind and SCANDI divergences from all clear sky data from 2008 to 2014 with separate data points representing the relationship within a two hour sliding window. The top histogram show the correlation coefficient of all data; the middle histogram shows the coefficient of determination of the data with a positive correlation; and the bottom histogram shows the apparent scale height estimations from the positively correlated data with at least a moderate strength relationship. This excludes a total of 61% of the original data-set.

heights from 2008-2014 winters at Svalbard at 250km altitude to suit the dataset used. The temperature range found was 633< T<1014 to give 30< H<50km scale heights, smaller than expected. Within this scale height range there were seven strong positive correlations out of a total of 5256 calculated correlations. This indicates a 0.1% chance of the vertical and horizontal wind being related by a realistic scale height and the full Burnside relation being satisfied.

A Gaussian is fitted to the data (blue line) to give a best estimate of the apparent scale height, yielding a value of 281±4km ($H_{best}$). This is 5.6-9.4 times larger than the scale height calculated by MSIS. However this estimate is comparable to proportions calculated in previous studies of the Burnside relation which have been found to be up to five times larger (Crickmore, 1993; Guo and McEwen, 2003; Anderson, 2011).

## 6.5 Discussion (Limitations of the Burnside equation)

Clearly the Burnside relation is not a good model of the Svalbard thermospheric winds as measured by SCANDI and the FPI. Only 39% of the dataset shows at least a moderate relationship between the vertical and horizontal winds and, from these data, nearly all give a poor estimation of the equilibrium scale height. The average shows a scale height estimation 5.6 times too large. However the spread of the data is large, which shows, firstly, that this is an imprecise estimation, and, secondly, that the relationship between the vertical and horizontal wind is inconsistent and not straight forward to model. A similar situation was found when using sub-populations of the dataset split by vertical wind magnitude or Kp index. Figure 6.1 is thus a sufficient display of the typical relationship found to continue with the investigation, for present purposes.

According to Burnside (1981), the relation should hold if the medium is well behaved and the behaviour of the oxygen atoms is representative of the bulk flow. A close examination of the assumptions within the derivation is required in order to try and understand why it breaks down. The assumptions by Burnside et al (1981) within derivation of relation which we choose to test are:

1. The pressure level height is constant.

2. Horizontal wind velocity and its horizontal gradients are constant with altitude.

3. Hydrostatic equilibrium holds.

For the Burnside relation to be investigated with the dataset at hand, we are making one additional assumption of our own:

4. The vertical wind is constant over the whole field of view.

### 6.5.1 Assumption 1: The Rate of change of pressure at the emission height is negligible

This assumption is precarious in that there are many known sources of heat in the high latitude thermosphere which drive the geopotential pressure height to new altitudes, e.g., auroral Joule and particle heating (see section 1.3). A persistently varying pressure height is expected. However, whether it has a significant effect and is a substantial fraction of the wind magnitudes, needs to be looked at in more detail.

Burnside et al (1981) show, to first order, that the horizontal divergence is accompanied by a vertical wind and not a varying pressure gradient. Crickmore et al (1993) explain that the change in the height of the given pressure level is a negligible source of error by reasoning that if all the observed vertical wind was due to a thermal expansion of the atmosphere then in two hours a parcel of air would have to change its altitude by 3 scale heights, which is unrealistic.

We use the CMAT2 model to study the rate of change of the geopotential height. In figure 6.2 the geopotential height of CMAT2's 33rd pressure level (at the grid point closest to Svalbard) is plotted against MLT for a typical day in January (22/01/12) under three different solar and geomagnetic activity conditions: quiet (Kp=2-) and active (Kp=5-) solar minimum (F10.7=75) nights in black and blue, and a quiet solar maximum night (F10.7=175) in red. These parameters are forced within the model and are not the values from the actual day. As shown, there is a clear variation in pressure level height throughout the day for all three sets of conditions ranging over 208-257km (altitudes of 630nm emissions).

The solar maximum night's pressure level is at a higher altitude than the solar minimum case by approximately 25km due to the increased solar flux heating resulting in the thermal expansion of the atmosphere. It also has the greatest change in pressure level height over the course of the day with a range of ∼21km altitude covered. The two solar minimum nights have more modest ranges of ∼15km for the quiet night and ∼18km for the active night.

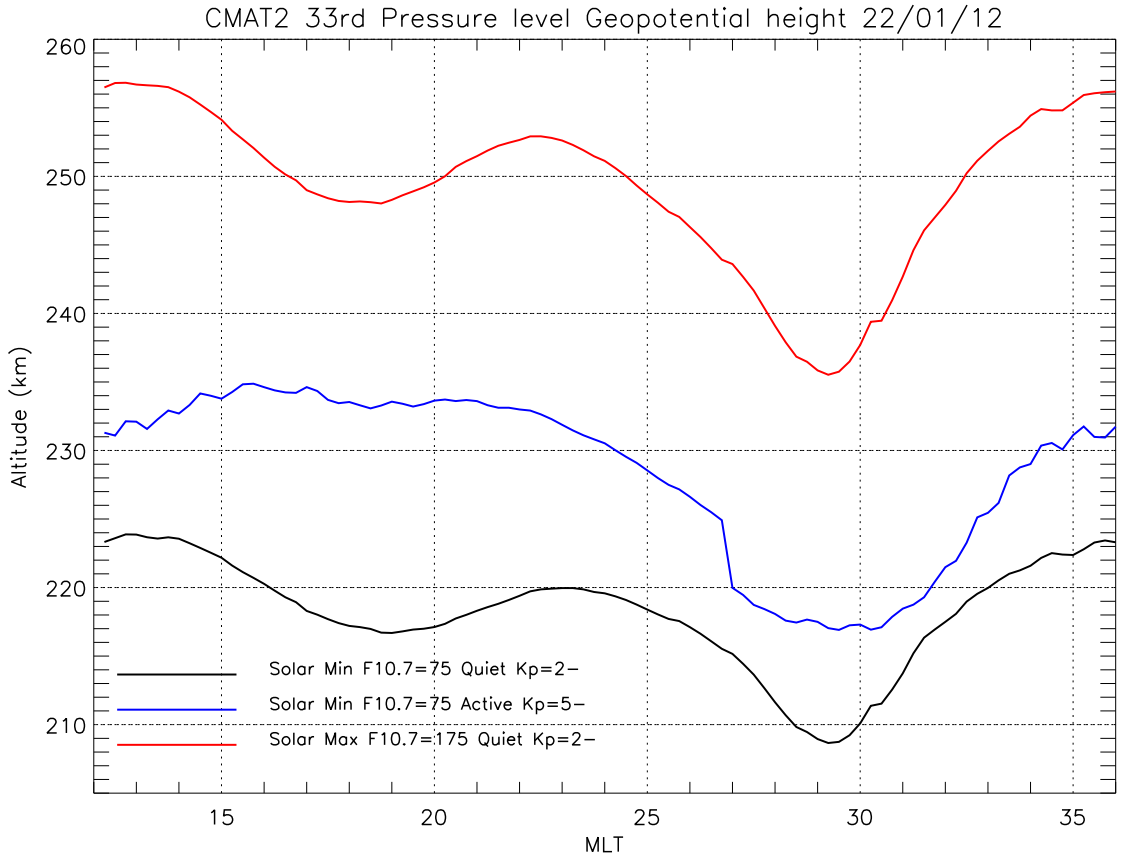


Figure 6.2: CMAT2's 33rd pressure level (five pressure levels down from top) geopotential height variation under three different combinations of solar and geomagnetic conditions. Data from the model's grid point closest to the Svalbard polar cap site.

The active night's pressure level is at a higher altitude than the quiet ones due to the increased particle and Joule heating. The increased particle precipitation is likely to be responsible for the relative variability and 'jerkiness' of this active night's pressure level height. This active night also exhibits the steepest drop in pressure level height of 5km in 12 minutes (26:48-27:00MLT). This is probably due to a relatively sharp relaxation of the atmosphere after a transit through the energetic magnetic midnight region. This complies with our findings from the previous chapter in figure 4.4 which shows that, under active conditions at midnight, the auroral oval expands overhead at Svalbard. However, for the other times of night, the auroral oval moves further equatorward away from the Svalbard site. As a result Svalbard experiences a corresponding sharp drop in geomagnetic heating which causes a sharp drop in geopotential height. The drop is calculated to be 7ms$^{-1}$ which is approximately the magnitude of the error bars of the vertical wind (5-10ms$^{-1}$). Therefore the CMAT2 simulations support the assumption that the rate of change of

pressure at the emission height is negligible.

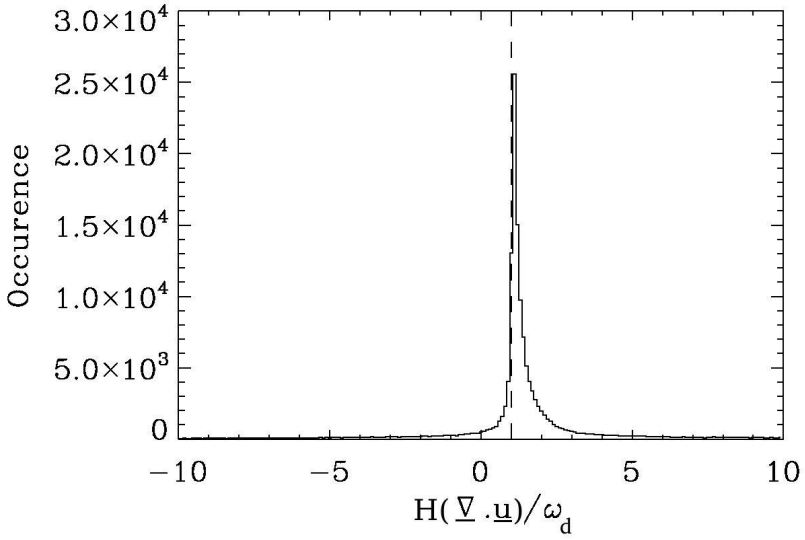


Figure 6.3: CMAT2's calculation of occurrence of different $\frac{H(\underline{\nabla}\cdot\underline{u})}{\omega_d}$ ratios over whole globe (where $\underline{u}$ is the horizontal wind, $\omega_d$ is the downdraft and H is the scale height) at the 33rd pressure level (a few down from the top pressure level). Set-up for a typical day with average Ap=6 and a moderate F10.7 value. The dashed line shows where ratio equals 1, i.e., the Burnside relation is satisfied. Courtesy of Timothy Spain (private communication).

Figure 6.3 is produced using the CMAT2 model data, for a geomagnetically quiet day with a moderate F10.7 value (Spain, private communication). It shows the global occurrence of the ratio $\frac{H(\underline{\nabla}.\underline{u})}{\omega_d}$ at the altitude 250km, where $\underline{u}$ is the horizontal wind, $\omega_d$ is the downdraft i.e. the vertical wind with respect to a pressure level, and H is the pressure scale height. This quantity should equal one where the Burnside equation applies, if we take $\omega_d$ to approximate the vertical wind velocity. Clearly, the graph shows a peak centred at a value of approximately one. This means that the Burnside relation works well in the CMAT2 model atmosphere, when considering the downdraft instead of the vertical wind. This makes sense as the downdraft is the vertical wind relative to a stationary pressure level, therefore it incorporates the constant pressure level height assumption into the parameter.

We now consider the implication of the downdraft on the FPI measurements. The FPI measures the absolute vertical wind $V_z$, which means that including the variation in the height of the pressure level, gives the vertical wind as:

$$U_z = \frac{dz_p}{dt} + \omega_d \tag{6.16}$$

where $\frac{dz_p}{dt}$ is the vertical speed of the pressure level. As shown in the CMAT2 model example of figure 6.2, the variation of the height of the pressure level is only a few $ms^{-1}$, well within the error bars of the FPI measurements. Therefore, the FPI measurements of absolute vertical winds should, within the typical error, be consistent with the downdraft values.

We conclude that the CMAT2 model atmosphere essentially does satisfy the Burnside relation in the rest frame of the pressure level, i.e., the constant-height pressure level assumption is a good approximation, which makes sense as Svalbard is a quasi-hydrostatic model. This compliance with the Burnside relation indicates that CMAT2's atmosphere must be well-behaved in other ways too, e.g., smooth temperatures and winds. This shall be explored in the discussion of the remaining assumptions that follow.

### 6.5.2 Assumption 2: Constant horizontal wind velocity with altitude

This is a reasonable assumption due to the large viscosity of the thermosphere which means that any small-scale vertical shears that arise are smoothed out very quickly. This is observed in the work of Killeen et al (1982) who analysed DE2 data to find little altitudinal structure between 200-400 km altitude. We investigate this using the CMAT2 model.

Figure 6.4 shows the variation of CMAT2's horizontal winds with pressure level and MLT under three different combinations of solar and geomagnetic activity conditions: Quiet (Kp=2-) solar minimum (F10.7=75) in top two plots and quiet solar maximum (F10.7=175) in middle two plots and active (Kp=5) solar minimum in the final two plots underneath. Typically the upper pressure level lies between 300-400km altitude. From figure 6.2 we have seen that the lower pressure level, 33, lies in the range 208-258km altitude. The colour scale is shown in the bottom plot with each interval corresponding to a difference of approximately $35ms^{-1}$. The general pattern of these graphs is what is expected as the Svalbard site moves through the pressure gradient-ionospheric convection pattern over a 24-hour period (see study in section 3.11.1).

The zonal winds under all three conditions are almost completely vertically stratified with very little altitudinal structure. The two quiet-time plots have very similar zonal wind structures with a maximum (i.e., most eastward) about 7 MLT and a minimum (i.e., most westward) about 15MLT. These are probably due to the return flow of plasma on the flanks of the ionospheric convection pattern imposing a general sunward movement

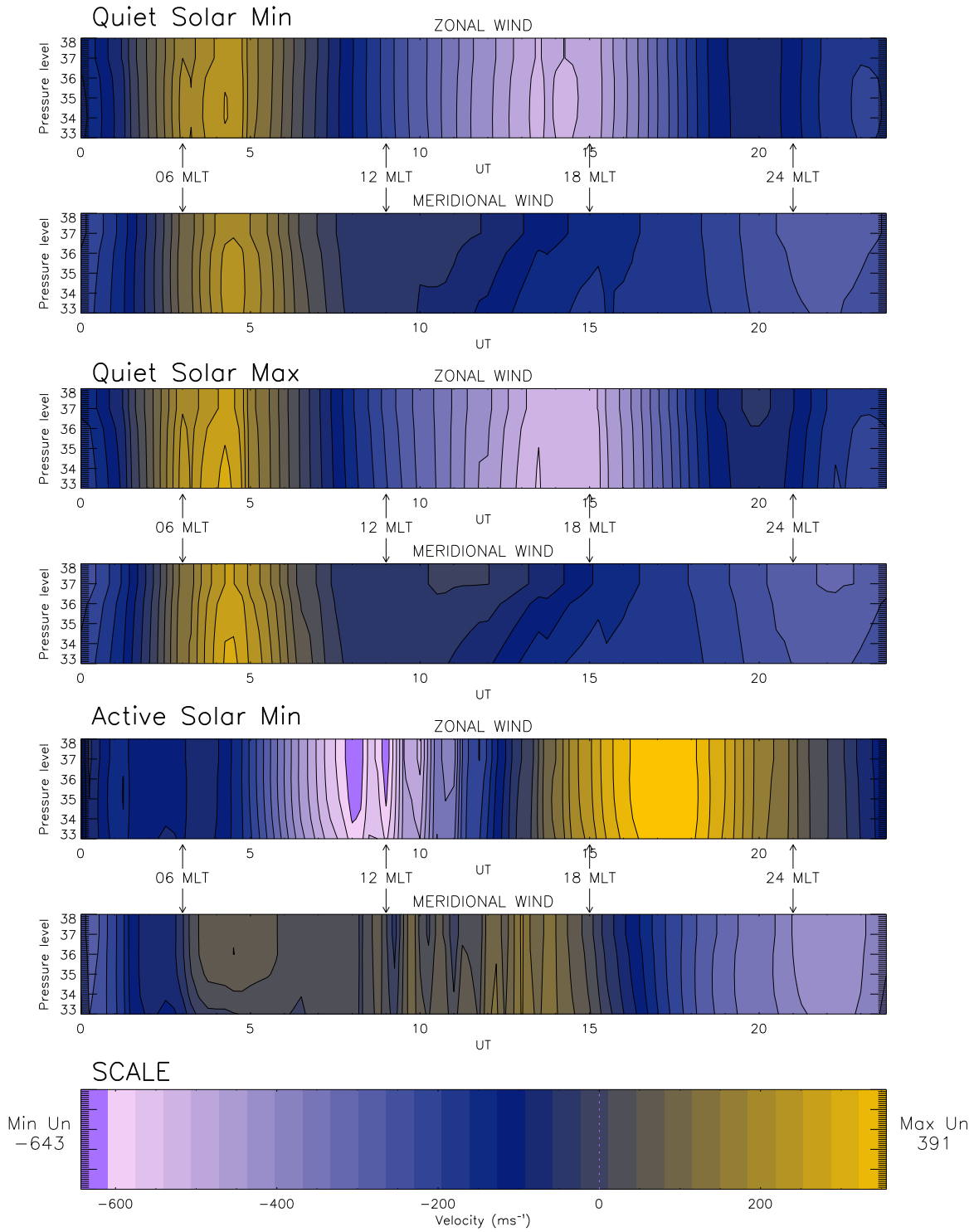


Figure 6.4: CMAT2 model zonal and meridional wind profiles for its 33rd to 38th pressure level under 3 different solar and geomagnetic conditions: Quiet (Kp=2-) solar minimum (F10.7=75) and solar maximum (F10.7=175) and active (Kp=5) solar minimum. Data are from CMAT2's grid point closest to the Svalbard polar cap site.

of the atmosphere via ion drag. The minimum at dusk is much more intense than the maximum at dawn probably due to the Coriolis force feeding in more plasma (and neutral gas) to the dusk-side cell, enhancing ion drag (and associated transfer of momentum) (Fuller-Rowell et al, 1987). The solar maximum zonal wind minimum period has a longer duration (∼3hours) and is more uniform compared to the solar minimum (∼1-2hours). The range in zonal wind speed is the same for quiet-time solar maximum and solar minimum (∼-600-300ms$^{-1}$) whereas for the active-time solar minimum this range is larger (-643-391ms$^{-1}$). The active-time zonal wind trend is very different showing a minimum (westward) about 9UT and a maximum (eastward) about 18UT, i.e., in the direction of the solar pressure gradient force. This is probably due to Svalbard's different relative position to the auroral oval. The active night zonal winds have the most altitudinal structure with vertical gradients evident from 7-12 UT, probably due to the enhanced day-side soft particle precipitation.

The meridional winds are less extreme than the zonal with fewer contour levels and a smaller magnitude range. Again the meridional winds quiet-time plots are similar with evidence of some altitudinal structure throughout, e.g., at 4:30UT and 10-16UT. The atmosphere is only vertically stratified for approximately two hours on either side of the structure at 7MLT. The active-time plot is very different, with persistent altitudinal structure throughout the night apart from the region from pre-magnetic-midnight sector up to approximately 5MLT. About 12-18MLT the meridional winds are very variable and show distinct wind shears and altitudinal structure. Nevertheless, this altitudinal structure still displays elements of vertical stratification. For example, there are constant long oval regions embedded in constant wind regions which create a vertical gradient, or there can be series of $\sim 45^o$ slanted areas of constant magnitude which have a component of vertical stratification.

For any UT where there is a significant altitudinal velocity gradient, where usually two ($\Delta$35ms$^{-1}$), very occasionally three ($\Delta$ 70ms$^{-1}$), contours are crossed. If we take an average difference change in altitude between the upper and lower pressure level to be 100km, this gives an approximate vertical gradient of horizontal wind to be 0.35-0.7s$\times 10^{-3} s^{-1}$. This would explain some of the break down of the Burnside relation but certainly not all. The predominant feature of these graphs is the vertical band-like structure. This shows that for the majority of the time showing there is little evidence of altitudinal structure. This model property validates the assumption of constant horizontal velocity with alti-

tude, thus supporting the Burnside relation, albeit adding some deviation at the particular times of strongest vertical velocity gradient.

### 6.5.3 Assumption 3: Hydrostatic equilibrium

Mostly the atmosphere is in hydrostatic equilibrium. However, this conditions is violated sometimes when particularly energetic events occur and the pressure level cannot adjust quickly enough. This is when Burnside-type behaviour will cease to exist.

The hydrostatic equilibrium equation can be thought of as the vertical momentum equation when there are only two significant terms present: the vertical pressure gradient and the acceleration due to gravity (see section 1.2.3 in chapter 1). This occurs when the change of pressure in the vertical direction is caused by the change of mass of overlying gas. If there are any accelerations, i.e., extra forces acting on the gas, then the atmosphere will be displaced from equilibrium. An example of this is when the rate of change of vertical velocity is large, a viscosity term becomes important in the vertical momentum equation so it becomes:

$$\frac{d\vec{U_z}}{dt} = \vec{g} - \frac{1}{\rho}\vec{\nabla}P - \frac{\mu}{\rho}\vec{\nabla}^2\vec{U_z} \tag{6.17}$$

where $U_z$ is the vertical wind. It can be seen in the above equation, the viscosity term is not important, then the equation reverts back to the equation of hydrostatic equilibrium (so long as there is no explicit time dependence of the vertical wind in the equation). Hence, in the case of rapidly varying vertical winds, hydrostatic equilibrium is violated.

Figure 6.5 shows a histogram of vertical wind accelerations for the Svalbard FPI 630nm data (2008-2014). Each acceleration data point is calculated by taking the difference in velocity between each pair of consecutive data-points and dividing by their time difference. Any accelerations shown are evidence that hydrostatic equilibrium breaks down. Nevertheless, most of the accelerations are negligible. The distribution peaks around zero and the main bulk of data lies between $\pm 0.5 \mathrm{ms}^{-2}$ where hydrostatic equilibrium is likely to be a good approximation. There are many examples where the acceleration reaches over $2 \mathrm{ms}^{-2}$ in which case it cannot be ignored and indicates there must be another significant term within the vertical momentum equation. Furthermore, this histogram has a slightly negative skew to the tail which shows there are more large-scale decelerations than accelerations in the thermosphere. This suggests that the extra term in the vertical

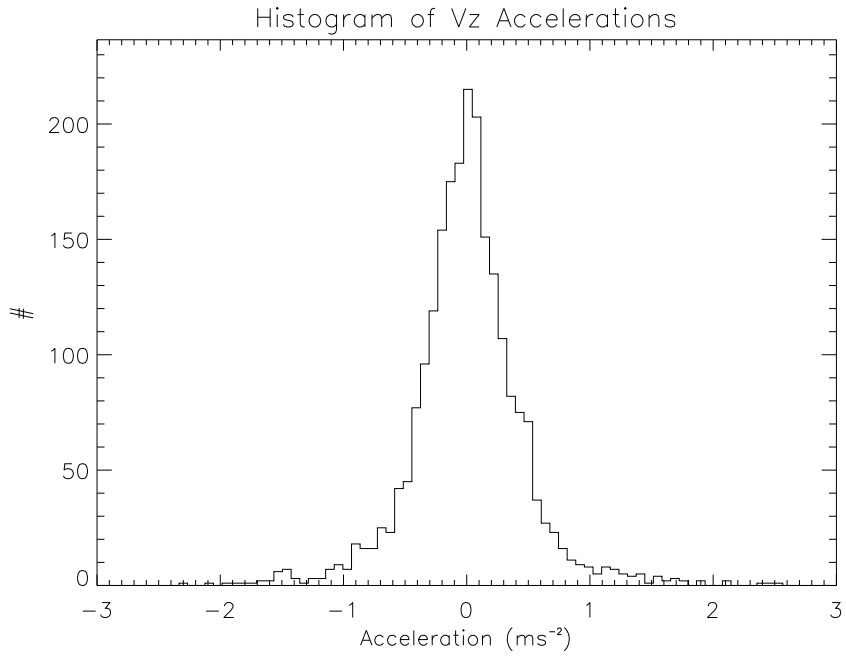


Figure 6.5: Histogram of vertical wind accelerations present in the Svalbard FPI data-set from 2008-2014.

momentum equation is viscosity which has a tendency to slow the upper atmosphere.

To further prove a violation of hydrostatic equilibrium, it is required to show that there is a vertical wind independent of the background pressure level movement. However, it is impossible to separate these with FPI observations. If there is a large heating event, there will also be a large vertical movement of the pressure level. Therefore, it is required to show that over the period of vertical motion there is a constant temperature, showing that the pressure level is stationary. This is assuming the continuity of particle density due to the dependence of pressure on density (P=nkT) which means that this method does not provide conclusive evidence but does provide strong indications that hydrostatic equilibrium is violated. There was not enough time to carry out this study but this must be investigated in the future. For now figure 6.5 has provided us with compelling indications that hydrostatic equilibrium does not always hold.

### 6.5.4 Assumption 4: Constant vertical wind over entire field of view

It is a commonly held assumption that the vertical wind is constant across the whole field of view. Previous research has shown that, at upper thermospheric altitudes, a typical horizontal spatial scale of vertical wind is ∼200km (see section 1.4.2) which is far smaller than SCANDI's field of view ∼1200km across (or the FPIs of ∼900km). Thus, this

assumption is a definite source of error. It would require multiple observing instruments to determine the vertical wind absolutely at multiple regions across the FOV and estimate the vertical wind scale size in the horizontal direction, as with the experiments by Anderson et al (2011).

## 6.6 Conclusion

The method by Burnside et al (1981) for calculating vertical winds from horizontal divergence has not proven suitable for our dataset. This does not necessarily mean that the relation is flawed; it could still be valid at low and mid-latitude observing sites; or it could be due to the limitations of our single site instruments prohibiting us from detecting the effect.

With a single observing instrument, it is impossible to resolve the vertical and horizontal component at each look direction absolutely. The line-of-sight observation are a combination of the two components, which has been shown to manifest as a Burnside type relation between the vertical and horizontal wind (see section 6.3.1). An attempt is made to resolve the two components of the wind by assuming that the vertical wind measured at the zenith look direction is constant across the FPI's entire field-of-view (FOV) (see section 6.3.2). This assumption is likely to be violated some of the time as the vertical wind horizontal scale size is shown to be highly variable (Spencer et al, 1982; Crickmore, 1993; Kosch et al, 2000; Price et al, 1995; and Anderson et al, 2001). If using two SDIs, common volume measurements could be made to infer the vertical and horizontal components with more confidence and a direct analysis of the relationship between the components could be carried out.

From addressing each assumption with the CMAT2 model simulations, we can see that the rate of change of pressure level height is unlikely to be a significant source of error. The assumption of constant horizontal wind with altitude above emission level could contribute to the discrepancy between the model and our observations, especially if the emission height is at lower altitudes than expected. However, this is unlikely under geomagnetically quiet and solar minimum conditions.

As a whole, CMAT2 is a well-behaved model which generally follows the Burnside behaviour. This is expected as it is a model which assumes vertical hydrostatic equilibrium. Therefore, we must be cautious drawing conclusions from the CMAT2 model about the

Burnside relation as this model is based on one of the assumptions on which the Burnside model is also based. This makes the Burnside relation seem like a better model than it actually is in the real atmosphere. Comparisons with a non-hydrostatic model such as GITM may be an interesting contrast to CMAT2 for this study. Although previous work has shown that most of the time, i.e., under geomagnetically quiet conditions, GITM has shown hydrostatic equilibrium to be a good assumption (Deng et al, 2007); GITM is the one of the few models that have been able to reproduce large vertical winds of up to 100 m/s.

At Svalbard, large scale vertical winds are sometimes present as well as occasional accelerations greater than $2\text{ms}^{-2}$, indicating occasional violation of hydrostatic equilibrium. For further evidence a study of the behaviour of the concurrent temperature behaviour with the large wind events needs to be carried out.

To conclude, the Burnside relation may still be appropriate for low-latitude sites but not for polar cap sites. The FPI community have held onto the relation for decades as it has the potential to be a very powerful tool, and they should not discard it yet; the low to mid-latitude observers should continue to use it with caution. The high-latitude sites, however, seem to have too many large vertical winds and heat sources that violate the assumptions within the relation. However, now with the high sensitivity of modern detectors, direct vertical wind measurements are more accurate than inferred ones, unlike at the time of Burnside et al (1981). This relation should only be used (at mid- and low-latitude sites) where direct measurements are not possible, e.g., at off-zenith look directions. A tristatic polar cap study needs to be carried out which can compare the distinct vertical and horizontal wind at one observing volume, in order to assess the impact of our instrument limitations on our ability to assess the validity of the Burnside relation.

This page was intentionally left blank

# Chapter 7

# Cusp

## 7.1 The Cusp Upwelling

### 7.1.1 Background reading/Motivation

The cusp is the region of the dayside magnetosphere/upper atmosphere where the pathway for the magnetosheath plasma is most direct, i.e., the flux of precipitating particles is high and the particles retain more of their original energy distribution's characteristics. It is situated at approximately 78° geomagnetic latitude and is only a few hundred kilometres across longitudinally. Smith and Lockwood (1996) provide a comprehensive summary on the cusp and related studies.

Most of the cusp-related research was done before the mid-nineties. However, in the last decade, with the recent development of instrumentation, there has been an injection of renewed interest, mainly due to several observations of density structures within and around the cusp region. Lühr et al (2005) provided evidence of a neutral upwelling within the cusp at high altitudes observing a density bulge at ∼400km altitude. This led Demars and Schunk (2007) to use the TGCM model to replicate this 'neutral fountain' effect by introducing an artificial ion-neutral heating. The Streak satellite found contrasting evidence of a density depletion at lower altitudes (Clemmens et al, 2008).

There are several potential physical mechanisms sustaining the thermospheric upflow: Joule heating in E region, soft particle precipitation heating, F region ion frictional heating (another way to consider Joule heating), solar EUV radiation and dissipating atmospheric

gravity waves (AGWs). AGWs are an unlikely candidate as they have a large diffusion coefficient so the majority of AGWs, sourced in the troposphere, dissipate before reaching the upper atmosphere. The minority that are produced in situ are not known to be persistent enough to maintain this structure. The solar EUV radiation is generally thought to be too weak to produce such a density enhancement; this is shown by Schlegal et al (2005) using the TGCM model.

Lühr et al (2005) reported the CHAMP satellite's observations at ∼400km, showing a doubling of the neutral density over many consecutive passes of the cusp region altitude. This study also showed the cusp to be a localised feature of the thermosphere of a few hundred kilometres across. The density is found to be enhanced over 1.2 times the background level ∼50% of the time and even more so with the GRACE satellite (Kervalishvili and Lühr, 2013). The upwelling feeding the bulge is indicated to be a continuous process lasting for at least several hours. The CHAMP satellite also found intense small scale FACs, with amplitudes of several hundred $\mu\mathrm{Am}^{-2}$ , accompanying the density enhancement and occurring independently of geomagnetic activity. This led Lühr et al (2005) to suggest localised Joule heating from below, such as from E-region ionospheric currents, driving the upwelling. This is also suggested by Schlegal et al (2005) who repeatedly observed cusp density structures using longer CHAMP datasets. They find density maxima to be, on average, at $75^o$ magnetic latitude between 10-11MLT with density minima found poleward, above $75^o$, on both the day and the nightside.

Demars and Schunk (2007) highlighted a problem with the mechanism being Joule heating in the E region. The quantity of energy introduced into the TGCM model in order to drive the upwelling of the neutral gas observed by CHAMP (Lühr et al, 2005) was unrealistically large. They increased their artificial ion-neutral frictional heating to 110 times a typical value in the cusp. However, Lühr et al's (2005) case study of a doubling of density was an extreme case. Schlegal et al (2005)'s more extensive study found the largest density values to be just 1.5 times greater than background density, which still requires an unrealistic quantity of energy deposited in the E region. Schlegel et al (2005) propose that the standard model Joule heating estimates are too small to sustain the continuous upwelling and that the inclusion of neutral winds might give a more realistic energy contribution.

The Streak satellite low altitude (123-325km) observations were very compelling as they contrasted with CHAMP's observations at high altitudes (400km). Clemmons et al

(2008) reported that Streak observed a weak density depletion of 1-2% within the cusp compared to adjacent areas. Streak's orbital height covers the altitudes of both E and F ionospheric layers and therefore this complies with the mechanism being Joule heating; potentially a combination from E region currents and F region ion-frictional drag. The density depletion could be due to the region of heating, which is losing mass from the upwelling neutrals, not being immediately replenished by the surrounding gas. Joule heating peaks at about 150km in the E region but can still be present in the F region and it is thus unclear which region is responsible for driving the upwellings.

Carlson et al (2012) propose that if energy is deposited in the F region where the atmosphere is tenuous and highly responsive, it is possible to generate the doubled density enhancement using a realistic source of energy. The cusp's characteristic soft particle precipitation could be responsible for directing the energy deposition to these high altitudes. Particle impact only provides a fraction of the energy required to heat the atmosphere sufficiently (Deng et al, 2013) - the dominant heat source is hypothesised to be the resultant ion-neutral frictional drag. Another characteristic of the cusp is the large electric fields generated by flux transfer events. These electric fields, combined with the soft particle precipitation, create significant Joule heating at high altitudes. This theory is corroborated by several modelling studies: Clemmons et al (2008) input soft particle precipitation in their model to generate vertical winds of $30\text{ms}^{-1}$ above 300km altitude, which are capable of generating a density bulge like that seen by CHAMP; Sadler et al (2012) incorporate FAST satellite particle measurement data into their model to generate vertical winds of $\sim 20\text{ms}^{-1}$, capable of producing a density enhancement of $\sim$1.04 times the background at 400km altitude; and Zhang et al (2012) introduce soft particle precipitation into the Coupled Magnetosphere-Ionosphere-Thermosphere (CTIM) model to generate a density enhancement of $\sim$1.25 times the background at 400km altitude. Deng et al (2013) drive their model multiple times, systematically introducing ion, electron and Poynting flux (representing Joule heating). In order to obtain density enhancements of $\sim$1.5 times in the cusp, soft particle heating needed to be combined with high-altitude Poynting flux. We must bear in mind that this model run was for equinox with a sunlit cusp; the experiment that follows involves observing the winter night-time cusp.

### 7.1.2 The CUSPN Campaign

In January 2012 the APL led the CUSPN campaign, in Longyearbyen, Svalbard, with an experiment designed to simultaneously and independently measure the charged and neutral components of the cusp region for the first time. The specific aim was to test the plausibility of the Carlson et al (2012) mechanism, i.e., find soft particle precipitation and evidence of F region ion frictional heating concurrent with the cusp upwelling. Presented are the first observations of this nature; concurrent, independent observations of the cusp's ionosphere and thermosphere, using the EISCAT Svalbard Radar (ESR) and UCLs FPIs with the support of the collocated Meridian Scanning Photometer (MSP).

The experiment was run on the morning of the 22nd January 2012 when a CME encountered the Earth, at 06:17UT, a couple of hours before the arrival of the cusp over Svalbard. This had the effect of energising the Dungey cycle, causing more intense flux transfer events (FTEs), creating ideal conditions to observe the cusp region. The arrival of the CME is evident in the IMAGE x-component magnetometer data, presented in figure 7.1, as an abrupt deflection of the geomagnetic field at 06:12UT which was felt as far south as Tartu, Estonia. Afterwards, between 08-10UT over the polar and auroral regions, particularly over Longyearbyen (LYR), there are pulses with periods of ∼200 seconds, known as Pc-5 ULF waves; good indicators of cusp precipitation (Anderson et al, 1993).

The ESR was used to measure the ionosphere, producing height profiles of the ionospheric parameters, i.e., the electron density, the ion temperature, the plasma velocity and the electron temperatures. The UCL FPIs measured the thermosphere. The narrow-field Svalbard FPI took high-time-resolution measurements of the zenith (cadence of 30 seconds) with intermittent observations of the north-east and south-west for horizontal wind measurements. The wide-field Scanning Doppler Imager (SCANDI), was used to monitor the mesoscale response of the thermosphere to cusp conditions across $\sim 10^o$ of latitude. This offered a large-scale thermospheric context for the cusp activity. From these we created mesoscale maps of the horizontal wind field, 630nm emission intensity and neutral temperature.

Figure 7.2 shows the comparison between the main datasets from 06-12UT. Subplot a) shows the electron density height profile observed with the 42m dish along the field-aligned position between 100-400km altitude with the colourbar to the right; subplot b) shows the

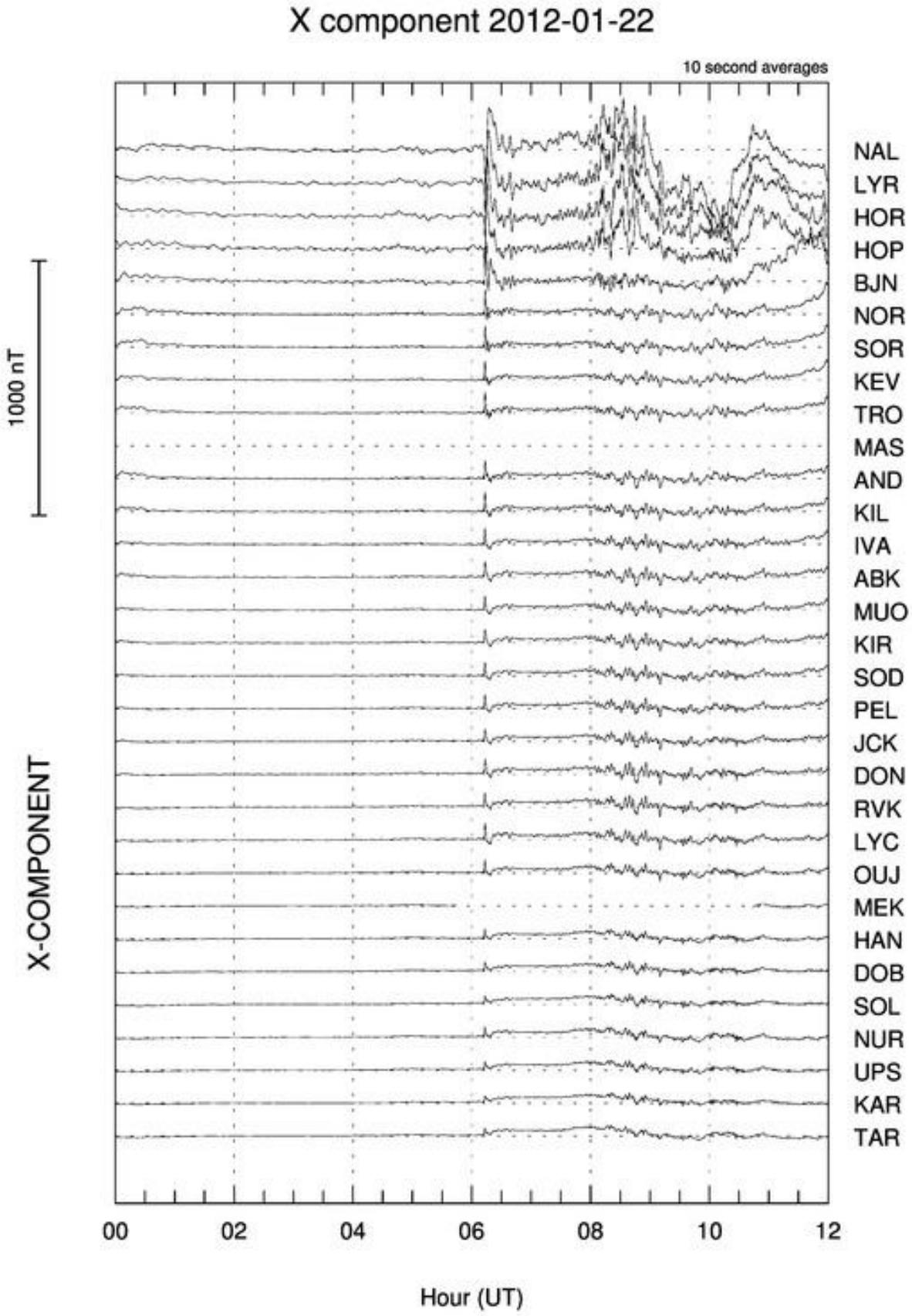


Figure 7.1: The IMAGE magnetometer array x-component of the geomagnetic field deflections from 00-12UT on the CUSPN night.

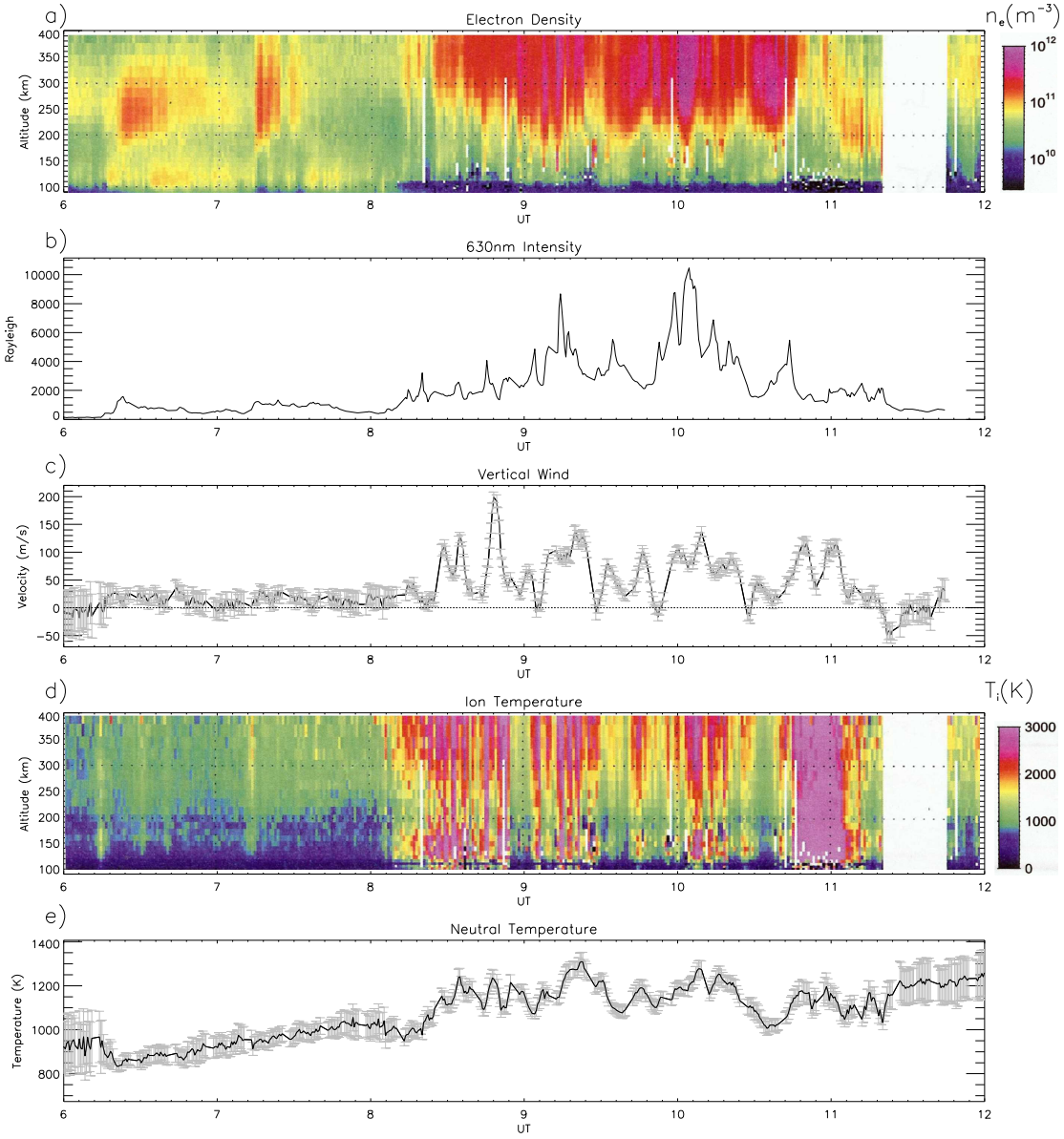


Figure 7.2: The main CUSPN data-sets from 06-12UT. Subplot a) shows the electron density height profile (100-400km altitude) measured by the ESR, subplot b) shows the zenith 630nm relative intensity measured by the FPI, subplot c) shows the zenith vertical wind measurements by the FPI, subplot d) shows the ion temperature height profile (100-400km altitude) measured by the ESR, and subplot e) shows the F region neutral temperature measured in the F region.

FPI zenith look direction's 630nm absolute intensity measurements (calibrated with MSP); subplot c) shows the corresponding zenith wind measurement representing the vertical wind with error bars in grey and the zero velocity line marked across in black. Subplot d) shows the ion temperature height profile in the same format as subplot a). Subplot e) shows the zenith neutral temperature measurements from the FPI corresponding to the data from b) and c). The FPI zenith look direction is vertically upwards whereas the ESR field-aligned look direction is at $181^o$ azimuth and $81.6^o$ zenith. Translated up to 250km altitude this results in the viewing volumes to be offset by ∼37km. With the large spatial scales of the the thermosphere it is safe to assume that the instruments are looking at the same viewing volume.

Considering the ESR data first in figure 7.2 a), the arrival of the CME is evident immediately in the electron density; as an intense pulse of F region particle precipitation at 06:20UT, followed by another one at 07:15UT. Both produced a density of $\sim 5 \times 10^{11} \mathrm{m}^{-3}$ between 200-300km altitude and there is also some residual E region precipitation reaching a density of $\sim 10^{11} \mathrm{m}^{-3}$ between 100-150km altitude. From 08:15-10:50UT there was a longer more sustained period of F region particle precipitation producing a maximum plasma density of $10^{12} \mathrm{m}^{-3}$; a magnitude usually only reached at daytime due to photo-ionisation. However, since this is a 24hr night-time, this cannot be the case here. As there is no concurrent E region particle precipitation, it seems likely that this is the cusp coming into view. The F region cusp precipitation causes the characteristic red glow of the cusp via electron impact excitation of atomic oxygen species followed by dissociative recombination. This is the dominant method of excitation at night-time when there is no photo-ionisation. Therefore we expect to see this swathe of plasma density enhancement seen in the ESR data accompanied by a red glow from the expected enhancement in atomic oxygen emission.

Using the UNIS MSP data it is possible to confirm that this soft precipitation is, in fact, cusp precipitation. The MSP can provide absolute estimates of the 630nm and 557.7nm emission intensities in Rayleighs. The cusp region is defined as the region where the 630nm intensity exceeds the 557.7nm intensity emission (Deehr et al, 1980). Figure 7.3 displays contour plots of the MSP data, courtesy of Dag Lorentzen (UNIS), for the night of the CUSPN campaign. For each plot the x-axes represent the UT ranging from 05:00-9:30UT, after which the MSP ceases its operation due to background light saturation. The y-axes represent the elevation angle of the instrument plotted as $180\text{-}0^o$ so that the plot is

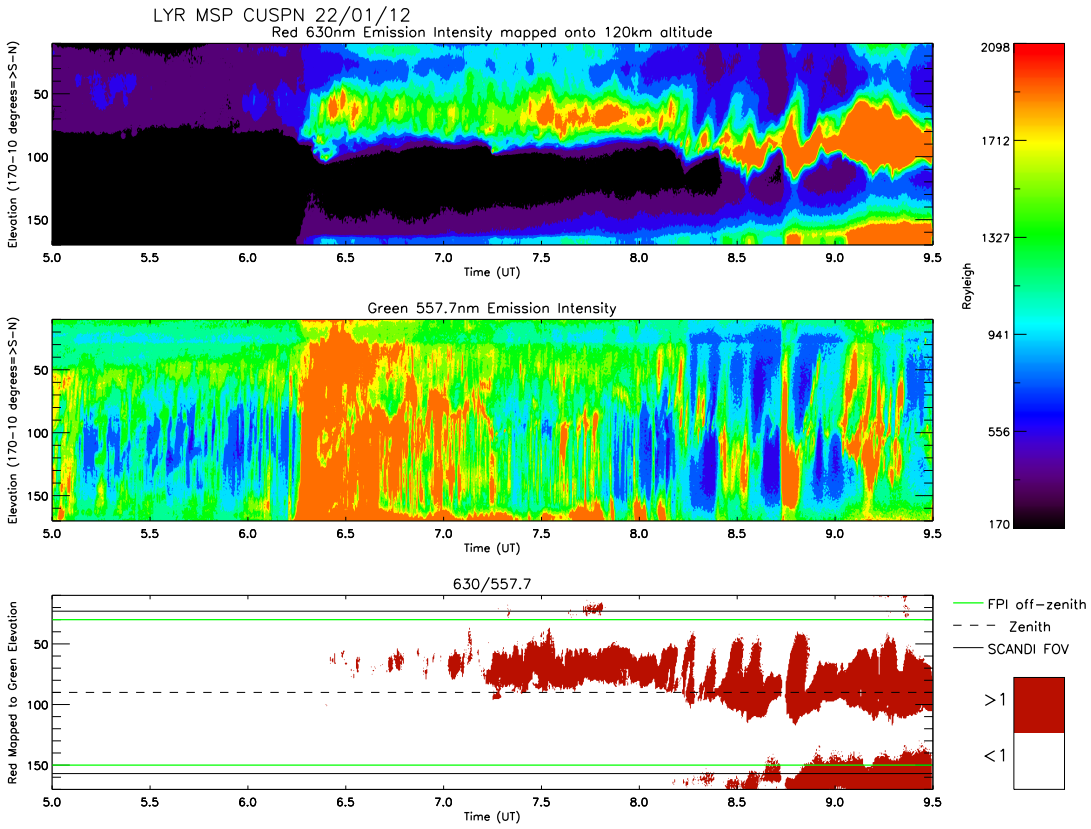


Figure 7.3: MSP data contour plots showing the absolute emission intensity of the red line emission (top), green line emission (middle) and the ratio between them (bottom) varying with UT on the x axis from 00:00-09:30UT, and elevation on the y axis from 180-0$^o$, i.e., south to north. The data (from the Kjell Henriksen Observatory (KHO) on Svalbard, Norway) are courtesy of Prof Dag Lorentzen at the University Centre in Svalbard.

orientated magnetic south to north. The top plot displays the 630nm emission intensity, the middle displays the 557.7nm emission and the bottom displays where the 630/557.7 nm intensity ratio exceeds unity, which is commonly used as a definition of the occurrence of cusp precipitation (Deehr et al, 1980). Overplotted are the edges of SCANDI's field-of-view as solid black lines, the off-zenith look directions for the FPI in green and the zenith look direction for both instruments as a black dashed line. Caution must be taken with data points at the edges of the plot as these are likely to be trees on the horizon contaminating the image. On the set of plots presented, the range of elevations regarded as reliable data are between 130-20$^o$.

First considering the top and middle plot in figure 7.3 we can see that before the CME arrived at the magnetosphere, the 630nm (red-line) intensities were relatively low ranging

from ∼170-700R; and the 557.7nm (green-line) intensities were relatively low too with values ranging between ∼700-1500R; the geomagnetic activity was quiet. The arrival of the CME can be seen at 06:20UT with a distinct increase in 630nm (red) intensities (F region precipitation) north of the zenith and across the entire magnetic meridian for the green line (E region precipitation). This behaviour continues and intensity strengthens for the red line until 09:30UT. However, the green line intensity decreases over this time (over interval ∼7:00-09:30UT). Looking at the bottom plot, the ratio of these intensities shows traces of cusp precipitation starting at 06:30UT. However, strongest cusp precipitation does not build up until 07:25UT which remains until at least 09:30UT at the end of the data-set. The strong cusp precipitation has a gradual equatorward drift meaning it only intersects the FPI zenith field of view (dashed line) at 08:25UT. This coincides with the arrival of the strong persistent F region particle precipitation evident in the ESR data; confirmation that this is a signature of cusp precipitation. From this analysis we can now be confident that both the ESR and the FPI are viewing the cusp region of the thermosphere.

Returning to figure 7.2 we may see the effect of the cusp precipitation on the thermosphere. Considering the FPI zenith 630nm intensities first, in subplot b), which have been cross-calibrated with the MSP data to give absolute values, we see that before the CME arrives the values are <100R, relatively small, representing little auroral activity; this was the case from 00-06 UT where $K_P$=[1-,1+]. At the arrival of the CME the intensities marginally increase hovering around 500R, however, it is not until the point of the cusp precipitation passing equatorward into the FPI zenith viewing volume when the intensities dramatically increase up to extreme values of 10.5kR. This increased intensity persists until the cusp precipitation passes out of the zenith field-of-view. All intensity values are bright enough not to be affected by the hydroxyl contaminating line (chapter 5) and so represent a reliable measure of the atomic oxygen red line emission.

Considering the vertical winds next, in figure 7.2 c), before the CME arrival at 06UT they look very quiet hovering about zero; the 12 hours previous are very much the same. Since this part of the night was so quiet, these data are used to calculate the baseline. When the CME hits, the vertical wind remains hovering close to zero but the error bars are much smaller due to the increased emission intensity. As the cusp precipitation comes into view at 08:25UT, the first upwelling begins reaching ∼110ms$^{-1}$ and lasting approximately 18 minutes. This is immediately followed by another one reaching a velocity of 200ms$^{-1}$

(the maximum velocity in the data-set), lasting 10 minutes. This is followed by a close succession of extreme upwellings, of 10-30 minutes duration, reaching average peak velocities of $\sim$90ms$^{-1}$ with the last ending at 11:06UT a short time after the main cusp precipitation passes out of the field-of-view ($\sim$10:50UT). The winds return to zero at 11:18UT. This amounts to $\sim$3 hours of an overall systematic upwelling averaging $\sim$50ms$^{-1}$; unprecedentedly large and persistent vertical winds. During this period the intensities are so large the error bars (median of 9.25ms$^{-1}$) are very small relative to the upwelling variations. Immediately after the upwelling period, there is a quick downwelling reaching -49ms$^{-1}$ at 11:25UT before returning to normal again at 11:30UT (hovering about zero). Due to the high cadence of FPI measurement, we can be confident in the reliability of the vertical wind data; each upwelling observation consists of multiple successive measurements, both on the increase and the decrease of the fluctuation. Furthermore, the individual spectra have been checked to confirm, by eye, the quality of each data-point.

Figure d) shows the ion temperature versus height profile (measured by the ESR) between 100-400km altitude. Before the cusp precipitation comes into view, the ambient ion temperatures are $\sim$1000K in the F region and a few hundred Kelvin in the E region. At the very start of the two pulses of particle precipitation, the ion temperatures increase to 1200-1800K for a few minutes in both E to F region altitudes; a very brief response to precipitation events which last 15-20 minutes. Just before the main cusp precipitation arrives overhead at 08:05UT, the ion temperatures increase up to $\sim$2000K in the upper F region. Shortly after, starting at 08:15UT, the ions in the E region (110-140km altitude) also reach $\sim$2000K; only accompanied by two very brief bursts of soft precipitation (at 08:13UT and 08:19UT). At the time of the cusp precipitation arriving, these ion temperatures increase dramatically to 3000K both in the E and F regions even though hard precipitation ($\sim 10^{11}$m$^{-3}$) produces only a tenth of the electron density produced by soft precipitation ($\sim 10^{12}$m$^{-3}$) and is far less persistent. The ion temperatures are highly variable throughout the entire time of cusp precipitation varying between 1000-3000K on time scales of 5-20 minutes. The ion temperatures return to their ambient values (1000K) just before the precipitation passes at 10:30UT. Once the cusp precipitation has stopped at 10:50UT, the strongest, most persistent and altitudinally uniform ion temperature enhancement occurs, reaching 3000K at times 10:45-11:05UT over 110-450km altitude, concurrent with only traces of F region precipitation. This accompanies a final double peaked upwelling reaching $\sim$130ms$^{-1}$.

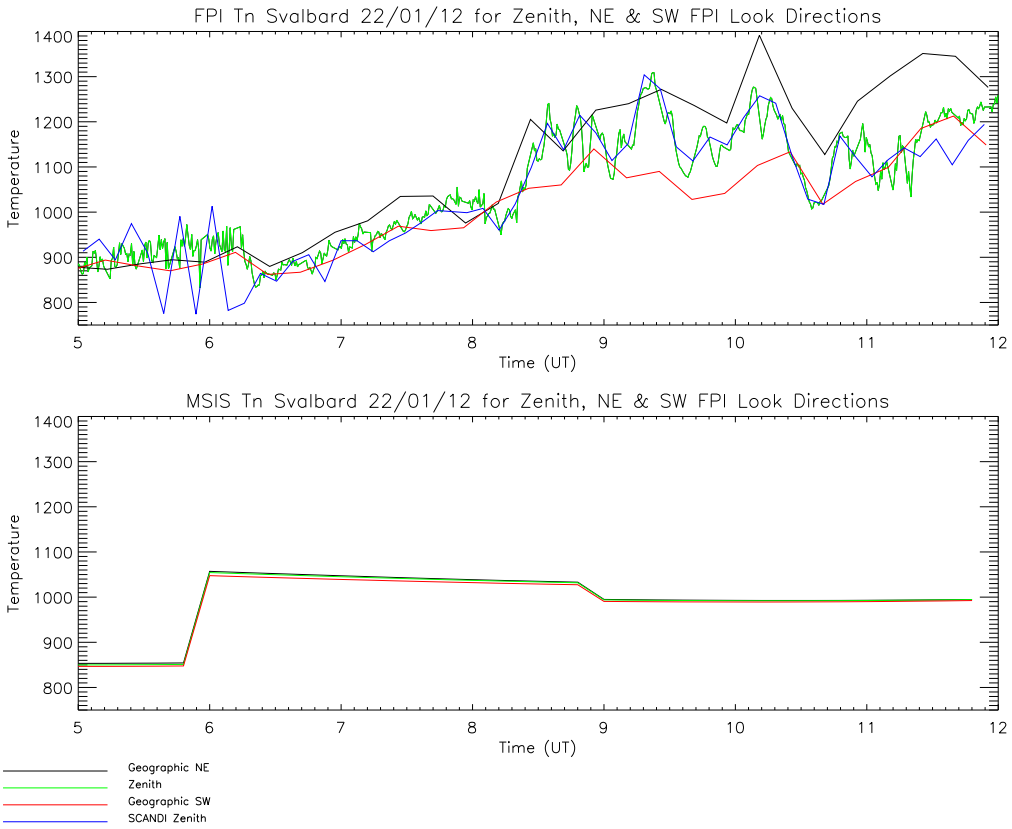


Figure 7.4: Svalbard temperature variation from 5-12UT on 22/01/12. Top plot shows real data with FPI zenith in green, SCANDI zenith in blue and FPI north-east and south-west look directions in black and red. Bottom plots shows MSIS data where green, black and red represent the same areas of the sky as on the top plot.

The variations in the neutral temperature are shown in figure 7.4 from 05:00-12:00UT. The top plot shows the neutral temperature variations measured by the narrow field FPI in the zenith (green), north-east (black) and south-west (red) viewing volumes with SCANDI's zenith viewing volume temperature variation overplotted in blue. The bottom plot shows the comparison with the corresponding data generated using the MSIS model with the FPI look directions' latitude and longitude used to find the temperature variations for the north-east and south-west look direction. First considering the FPI zenith, the temperature is ∼900-950K before the CME hits, however as soon as the CME arrives it drops to ∼850K. This is likely to be due to the onset of the E region precipitation driving the emission volume to a lower altitude. From 06:17UT to 08:00UT the temperature slowly ramps up to ∼1000K, drops slightly and then the cusp precipitation comes into view. At 08:15UT the zenith temperature increases dramatically to ∼1200K at 08:40UT,

after which the temperatures remain high, fluctuating between ∼1025-1325K, potentially due to ion-frictional heating and particle heating, until the end of the dataset at 12UT. This variation is consistent with the lower time resolution of the SCANDI zenith measurements (apart from before the CME arrival when the intensities are low and the SCANDI data are noisy) which gives confidence to each dataset. The other two look directions show a general north-east to south-west temperature gradient which becomes more pronounced at the time the cusp is overhead. It is a concern that the large temperatures to the north could be a result of hydroxyl contamination where the intensities are low.

The MSIS plot at the bottom of figure 7.4 shows a poor match to the observations highlighting how poor MSIS is at modelling the cusp region, particularly at CME times. This is likely to be due to the crude 3-hourly Kp index input as a proxy for geomagnetic activity and the limited spatial resolution.

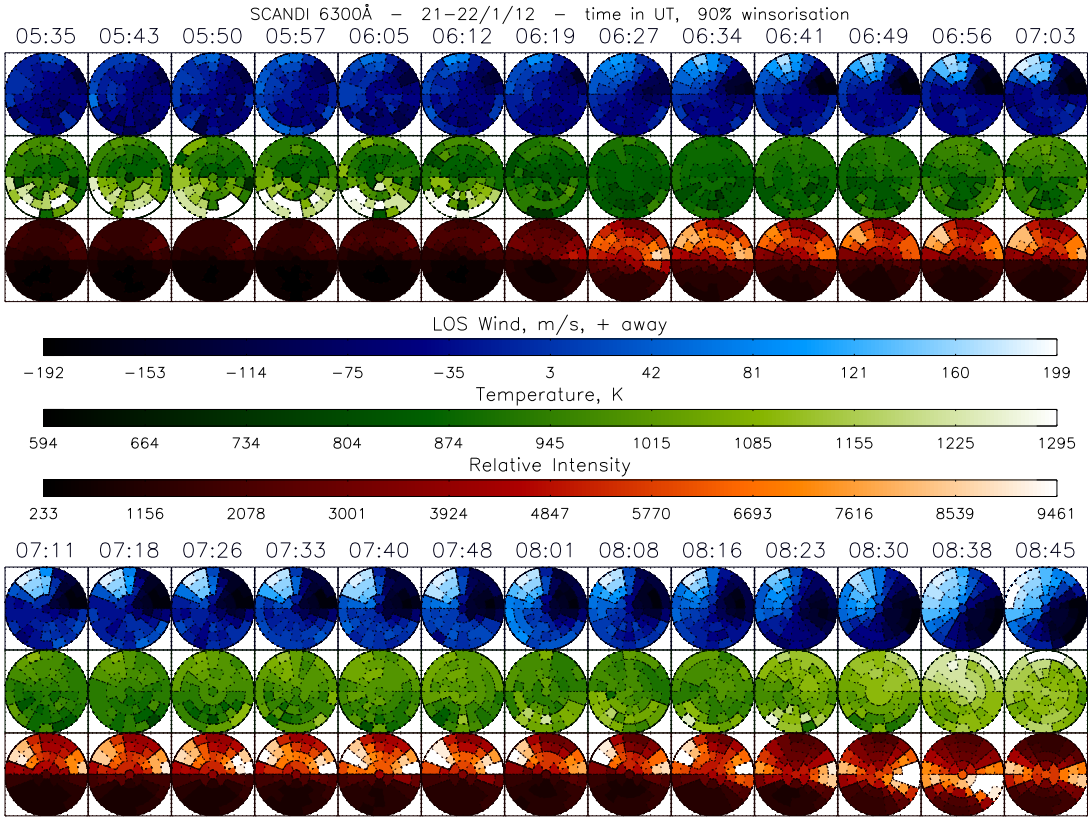


Figure 7.5: 61 zone map plots from 05:35-08:45UT on 22/01/12. Top row is line of sight neutral winds, middle row is the neutral temperature and bottom row is relative intensity of 630nm red O emission

Figure 7.5 and 7.6 show the all-sky SCANDI data as a set of 61 zone maps of a circular region of the 250km altitude thermosphere of diameter 1076km, centred above Svalbard.

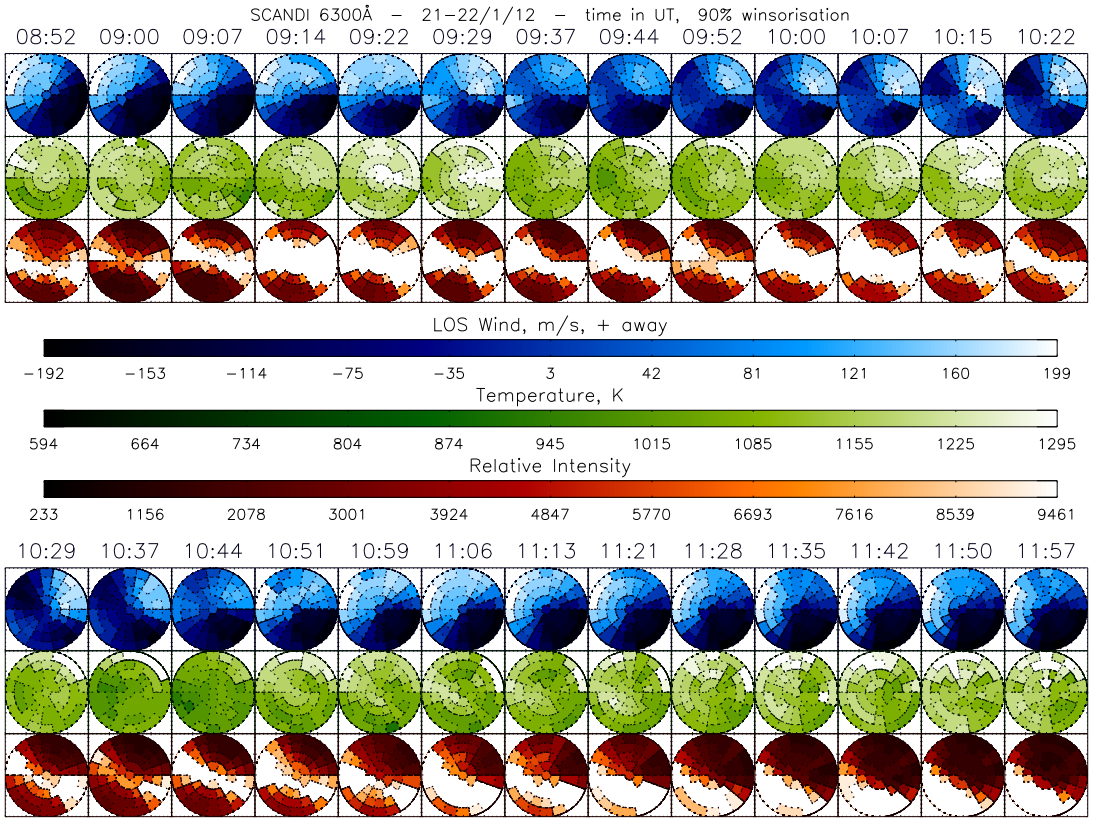


Figure 7.6: 61 zone map plots from 08:52-11:57UT on 22/01/12. Top row is line of sight neutral winds, middle row is the neutral temperature and bottom row is relative intensity of 630nm red O emission

The field of view is wide enough that it is possible to image the entire cusp region (a few 100km across). However, with each zone spanning ∼100km across it is also possible to resolve mesoscale features of the region. The top rows of plots in blue show the line-of-sight wind, the middle plots in green show the neutral temperature and the bottom plots in red show the relative intensity of the 630nm emission.

First considering the 630nm intensity map plots (red) in figure 7.5, there is a distinct difference in the behaviour of the atmosphere from before and after the CME arrival. Before, 05:35-06:12UT, low intensities (∼200-2000 counts) are uniformly spread across the field of view. Afterwards, the intensities are significantly higher to the magnetic north (∼4000-8000 counts) however remain low to the magnetic south (200-2000 counts). This is attributable to the slow onset of the cusp precipitation which commences at ∼06:30UT to the north of zenith as seen in the MSP data (figure 7.3). The 630nm emission to the north intensifies and slowly migrates equatorward until it reaches overhead the zenith at 08:23UT. This is consistent with the MSP data. Therefore, we can have some confidence

that the bright band in the map plots is the cusp passing overhead. This SCANDI dataset lasts longer than that of the MSP for the relevant event's time so it is possible to track the movement of the cusp throughout the whole experiment.

The intensities in figure 7.6 show the cusp remaining overhead and slowly drifting equatorward, corresponding with the MSP data. The region of the cusp precipitation reaches its full strength throughout these time-stamps and this is reflected in the intensity uniformly reaching the maximum value of 9461 counts within the cusp region. At some time stamps, the maximum 630nm intensities spread to cover more of SCANDI's zones, e.g., at 09:22UT, and at the next time stamp this area contracts and later expands in a pulsating fashion. This occurs to the end of the dataset (11:57UT), indicating FTEs. The illuminated region typically spans 3-4 of SCANDI's zones north-south, showing that the cusp stretches ∼300-400km latitudinally (∼3-4° latitude). The cusp region drifts out of the zenith zone at 10:51UT at the same time the signature of the cusp precipitation stops in the ESR data (figure 7.2 a)). The vertical wind, however, continues upwelling and does not return to zero until ∼11:20UT. At all other times of increased zenith zone intensity, from 08:48-10:51UT, there is a strong vertical wind velocity (averaging ∼90ms$^{-1}$). This is consistent with the narrow-field FPI intensities (figure 7.2 b).

Next considering the temperature variations in green (middle row), before the CME arrival there was a temperature gradient with the equatorward side of the field of view (∼1300K) hotter than the poleward side (∼950K). As soon as the CME hit, the whole of SCANDI's FOV dropped in temperature to ∼800-950K. This is due to the period of E region precipitation between 06:30-07:15UT causing the emitting layer to drop to a lower altitude in the thermosphere where it is cooler. After 07:18UT the temperatures gradually rise to more typical F region temperatures, and continue rising throughout the entire period of cusp precipitation up to values of 1295K. In this period, there is also a north-east to south-west gradient in temperatures consistent with the FPI.

Figure 7.5-7.6 also shows the line-of-sight (LOS) wind data in blue (top row), with positive values representing a red Doppler shift. These are the raw wind data, from which, a horizontal wind field is fitted (see section 3.11) which is much easier to interpret. The fitted horizontal wind fields are presented in figure 7.7 showing data from 05:50-11:57UT, formatted as circular fields of view of 1076km diameter at 250km altitude with a wind vector corresponding to, and positioned at, each of SCANDI's zones. The variations seen in this dataset comply with those from the raw LOS data (blue in the previous map plots)

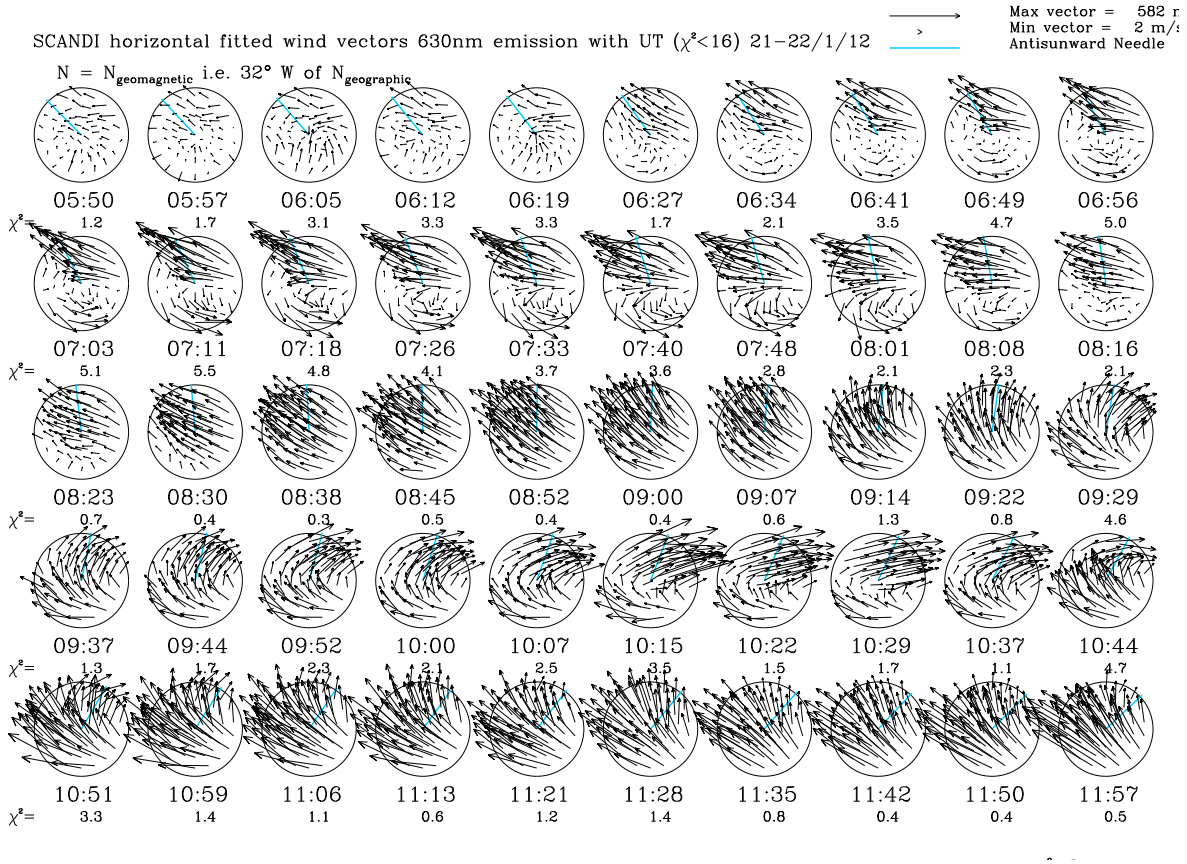


Figure 7.7: SCANDI's fitted horizontal wind vectors calculated from the 630nm emission line and plotted for 50 consecutive exposures on the morning of the 22nd January 2012 i.e. CUSPN campaign.

demonstrating the wind-fitting procedure was successful.

Evident in the wind-fit plots is a gradual response to the arrival of the CME, typical of the thermosphere due to its high inertia.

Before the arrival of the CME, 05:50-06:12UT, there is a shear in the wind field with the wind in the southern half of the FOV slowing mainly north-eastward and being small magnitude ($\sim$100ms$^{-1}$) and in the northern half flowing westward and being generally stronger magntiude (70-210ms$^{-1}$). It is not until ten minutes after the arrival of the CME, 06:27UT, that the winds look noticeably different. The wind pattern remains the same (a vague north-south divide) and the south side winds remain small, however, the north side strong westward winds have spread further south slightly, strengthening to become more uniformly $\sim$200ms$^{-1}$. The flow at the northern side of the FOV continues to strengthen and slowly engulfs increasingly more of the FOV in the southward direction, until at 08:38UT, the entire FOV is uniformly flowing in the same north-westward direction at a large magnitude of $\sim$210ms$^{-1}$.

This area of strong north-westward flow, since the arrival of the CME, has exactly coincided spatially with the area of the increased 630nm intensities seen in the map-plot. Furthermore, this north-westward flow direction is consistent with the solar pressure gradient force direction at 06:19-08:37UT (magnetic time 09:19-11:38MLT). The solar pressure gradient force direction at Svalbard is demonstrated in figure 3.13 in the climatological wind study in section 3.11.1 in chapter 3.

At 08:30UT the cusp region is fully strengthened (figure 7.3) and lying overhead, slightly to the north (and ten minutes after the arrival of cusp precipitation in the zenith zone), the winds remain in the roughly uniform strong north-westward wind, consistent with the solar pressure gradient force. This persists until 09:14UT when the winds turn to a mainly north-eastward direction in the northern half poleward sector of the plot in line with the antisunward direction (ion-drag indication). This shows a possible lag time of ∼45minutes for the thermospheric winds to respond to the cusp precipitation. Meanwhile (from 08:30-09:14UT), the southern section (equatorward at noon) remains north-westward; an indication of pressure gradient force. This north-south divide creates a large shear in the neutral wind. The northern half of the FOV's north-eastward winds reach their maximum value of $\sim 580 ms^{-1}$; indicating the relative strength of the ion drag force at this time. This configuration remains until 10:51UT when the winds return to a more uniform north-westward direction with a mild shear (however still large magnitude winds); seemingly returning to being controlled by the solar pressure gradient force. At 10:51UT (and up until the end of the observations 11:57UT) the vectors at the bottom of the FOV show maximum wind magnitudes of $\sim 580 ms^{-1}$. This indicates that as the time of maximum solar presure gradient forcing is approached (for Svalbard at 14UT) the solar pressure gradient force can grow strong enough to produce as large wind magnitudes as the ion-drag force seen earlier (e.g., 10:15-10:37UT).

## 7.2 Discussion

At the time of arrival of the cusp precipitation, there is an immediate response in all data-sets: the 630nm emissions intensify, the vertical winds upwell, the horizontal winds strengthen (in $-\nabla P$ direction which is poleward around noon), the neutral temperatures increase, the ESR ion temperatures increase and there are large deflections in the x-component of the IMAGE geomagnetic field. These parameters are highly correlated

throughout the entire ∼2.5 hours of cusp precipitation. The ion temperatures are not completely correlated as they are found to increase a little before the onset of cusp precipitation into the zenith FOV, ∼10 minutes. The ion temperatures also dramatically increase immediately after the period of cusp precipitation passes out of the zenith FOV, lasting persistently for 20 minutes. The plausibility of F region ion-frictional heating actuated by soft particle precipitation driving the cusp upwelling, and feeding the cusp density bulge, is discussed in this section.

### 7.2.1 Ion-frictional Heating

#### Criteria

In order to conclusively identify an ion-frictional heating event, i.e., when the collisional transfer of momentum to heat between ions and neutrals causing both to rise in temperature, two criteria must be satisfied:

1. the ion temperatures must be raised above the neutral temperatures by 100-200K (McCrea et al, 1991; Maurice and Hanson, 1982)

2. the ion velocities must exceed the neutral velocities by an amount sufficient to allow heating by ion-neutral collisions and such that the ion temperature increase is difficult to explain by any other means, e.g., heat conduction or heat exchange with electrons.

#### Ion Energy Equation

In the F region on time scales greater than a few seconds and spatial scales of >15m the ion energy equation simplifies to (Maurice and Hanson, 1982):

$$T_i - T_n = \frac{m_n}{3k_b}(v_i - v_n)^2 \tag{7.1}$$

where $m_n$ is the mean molecular mass of neutrals, $k_b$ is the Boltzmann constant and $v_i$ and $v_n$ are the ion and neutral bulk flow velocities. This describes an atmosphere where there is a direct balance between the ion-neutral frictional heating (right-hand-side term) and thermal heat exchange with the neutral gas (left-hand-side term); it demonstrates the process of F region Joule heating. Above, many terms in the full ion energy equation have been neglected, but the main ones are as follows:

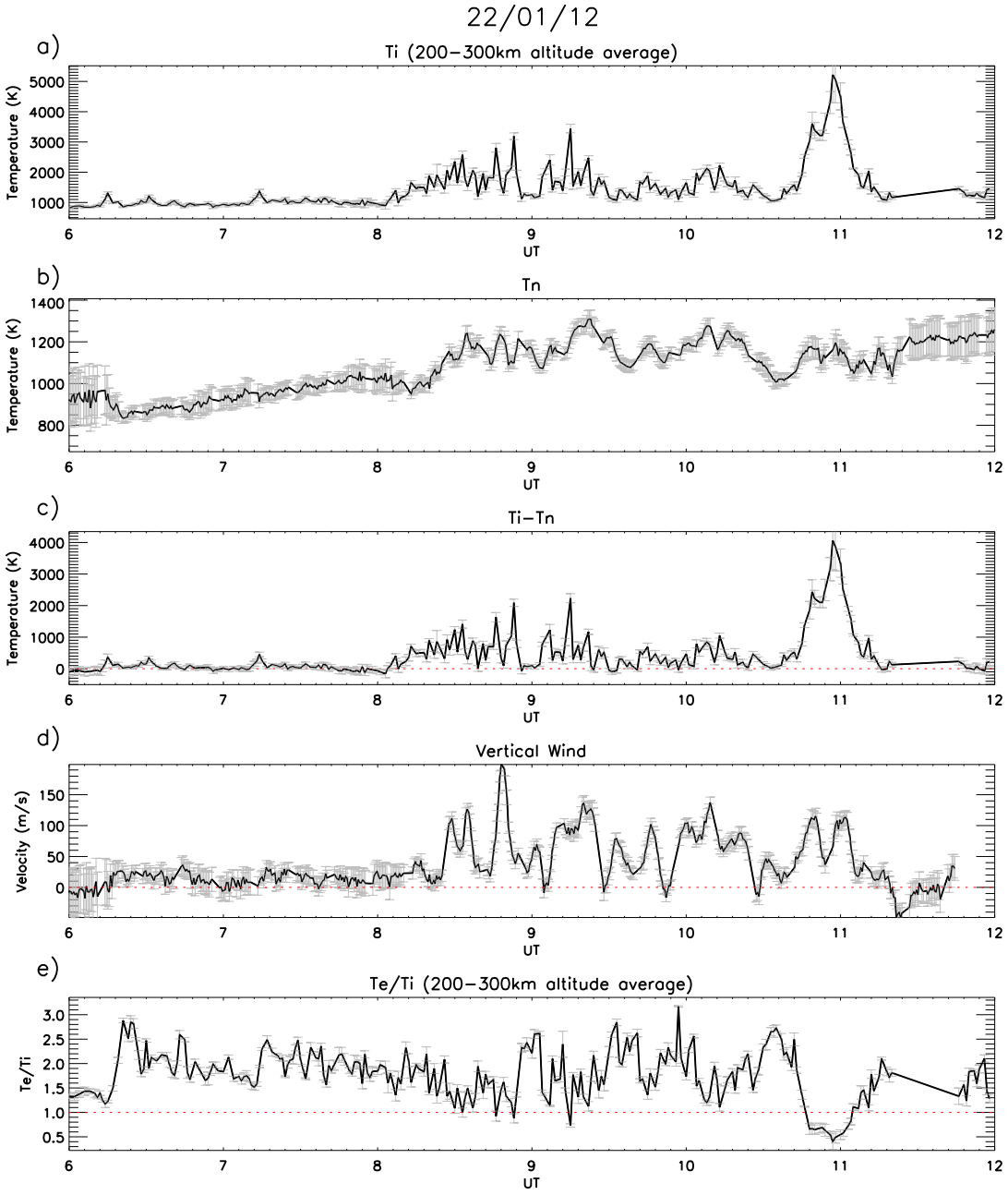


Figure 7.8: The CUSPN temperature data with the vertical wind data variations from 06-12UT. Subplot a) shows the ion temperature averaged between 200-300km altitude, subplot b) shows the zenith neutral temperature measured by the FPI, subplot c) shows $\|T_i\text{-}T_n\|$, subplot d) shows the vertical wind measured by the FPI and subplot e) shows the ratio between the electron and ion temperature.

1. Frictional heat exchange with the electron gas ($\propto |v_i - v_e|^2$) since this is only important in the E region, where there are significant differences in the ion and electron drifts in the direction parallel to the field line. The ions and electrons in the F region plasma largely drift together.

2. Thermal heat exchange with the electrons ($\propto |Ti - Te|$) as this increases ion temperature by <200K (Schunk and Sojka, 1982) which is dwarfed by typical ion-frictional heating rates.

3. The heat advection and heat conduction terms as they have been shown to be insignificant in the ion-energy equation at high latitudes, up to ∼400km altitude, by Schunk and Sojka (1982) and will also be dwarfed by typical ion-frictional heating rates.

Maurice and Hanson (1982) estimate an ion-neutral velocity difference of 400-600ms$^{-1}$ in order to achieve a temperature difference of 200K in the F region, and, McCrea et al (1991) estimate a difference of 1000ms$^{-1}$ in order to achieve a difference of 900K at 300km altitude. Maurice and Hanson (1982) present several case-studies of ion-frictional heating all showing strong shears in ion-drift with velocities reaching 1000-2000ms$^{-1}$ accompanied by ion temperatures reaching 2000-4000K. Every one of their case studies shows a strong correlation between the magnitude of the ion drift and the measured ion temperature.

**Other Ion-frictional Heating Indicators**

There are other indicators of the presence of ion-frictional heating as follows:

- Shears in ion velocities are expected due to the ions being accelerated above the neutral velocities. After a time lag of 0.5-6.5hr (Kosch et al, 2001) the neutrals may also experience a shear in velocity as the ion drag overcomes the high viscosity of the neutral atmosphere, however, usually the ion drift fluctuations are too short-lived to have an effect. As soon as the neutral wind becomes influenced by the ion-drag the ion-frictional heating will reduce as there are not as many ion-neutral collisions (Maurice and Hanson, 1982).

- Neutral temperature increases are expected, as the neutral heating rate is similar to that of the ions. However, the increase in neutral temperature will not be as large

as for the ions due to the relatively large heat capacity of the neutral gas, hence an ion-neutral temperature difference emerges.

- In extreme events, the temperature of the ions will exceed that of the electrons. The electron temperatures are almost always at a higher temperature than the ions (usually $T_e > T_i > T_n$). However, the frictional heating rate of the electrons is smaller than for the ions due to their fewer collisions per unit time with the neutrals. Therefore, in extreme events, the ion temperature can sometimes be elevated above that of the electrons.

**Optimum Altitude of Ion Frictional Heating**

The optimum altitude of Joule heating to actuate the cusp upwelling is a point of contention. Ion frictional heating is strongest at ∼150km altitude. However energy deposited at this altitude alone may not be sufficient to lift the dense air above. It is easiest to lift the atmosphere high in the F region where the atmosphere is the most rarefied; however at a location too high there will be insufficient mass in the column above to create the doubling of densities observed by satellites. It is possible that there may need to be a joint application of Ohmic E region Joule heating (from currents) and F region ion-frictional Joule heating. Ohmic and ion-frictional Joule heating describe the same phenomena, however, usually Ohmic is used to describe E region Joule heating (due to this being the region of currents), and ion-frictional is used to describe the F region Joule heating (see section 1.3.2 in chapter 1). This convention will be used from here-on in this study.

### 7.2.2 Identifying Ion Frictional Heating in the CUSPN Experiment

Unfortunately there is no data for the horizontal ion velocity as, during the CUSPN campaign, before our experiment, the driver motor for the 32m ESR dish broke due to severe storms so we could not perform beam-swinging. Without horizontal ion velocities, we could not calculate Joule heating energy deposition height profiles as desired. In addition, we could not conclusively confirm the existence of ion-frictional heating. Nevertheless, with our ion and neutral temperature data, we can make strong inferences about its existence. Since there is such strong correlation between $\|T_i\text{-}T_n\|$ and $\|v_i\text{-}v_n\|$ at high ion temperatures (2000-4000K (Maurice and Hanson, 1982)), this is a reasonable proxy for $\|v_i\text{-}v_n\|$. A minimum requirement for the ion-neutral temperature difference is 200K so

that it exceeds any possible electron heating (<200K, Schunk and Sojka (1982)) and we can negate this as a possible heat source and be more confident we are observing ion-frictional heating. We are expecting greater temperature differences than 200K by ion-frictional heating judging by the extra energy supplied by the CME.

Figure 7.8 shows the CUSPN experiment temperature data with the vertical wind from 06-12UT. Subplot a) shows the ion temperature variation averaged over the whole F region (200-300km altitude), b) shows the neutral temperature variation, c) shows the ion-neutral difference in temperature, d) shows the vertical wind variation, and e) shows the electron-ion temperature ratio, $T_r$.

Averaging the ion temperatures over the upper F region allows us to consider potential ion-frictional heating events occurring at the altitudes of cusp precipitation. These temperatures start to rise at 08:02UT whereas the neutral temperatures start increasing as the cusp precipitation arrives overhead at ∼08:15UT. This could be due to the cusp lying directly northward of the FPI's zenith, altering the electric field adjacent to it and causing a shear in the ion velocities above Svalbard, which would raise the ion temperatures but not the neutrals.

Throughout the period of upwelling, from 08:25-11:20UT, the F region ion temperatures reach ∼1500-5000K whereas the neutral temperatures reach ∼1000-1300K; the differences between them reach ∼200-4000K; all differences are larger than 200K, indicating potential ion-frictional heating. The average neutral temperature is ∼1200K and the average ion temperature is ∼2000K, this gives an average ion-neutral velocity difference of 1000ms$^{-1}$ according to the ion energy equation.

Throughout the period of cusp precipitation, from 8:25-10:15UT the $\|T_i\text{-}T_n\|$ values range between 200-2200K and the trend generally correlates with the upwellings. The two largest differences in this interval at 08:55UT and 09:15UT, are roughly 2200K, however they do not produce equal magnitude upwellings. The first event produces the largest upwelling reaching 200ms$^{-1}$ and the second is 140ms$^{-1}$ in magnitude. The subsequent upwellings do not have a proportional magnitude to the corresponding $\|T_i\text{-}T_n\|$ (if we use the first two upwellings as an indication). Evidently, the heating effect is not linear with F region heating. This indicates heat sources at lower altitudes contributing to the 'uplift'.

Throughout the cusp precipitation period, $T_r$ is mostly greater than 1 (electrons hotter than ions), apart from three occasions where it dips below 1 (ions hotter than electrons) for 1-2 minutes at 08:50UT, 08:55UT and 09:15UT. The first two times sandwich the largest

upwelling of the dataset (reaching 200ms$^{-1}$) and the third occurs at the start of the double peaked upwelling reaching 140ms$^{-1}$ occurring at 09:15UT. This provides strong evidence that ion-frictional heating is active in the upper F region.

The magnitudes of $\|T_i\text{-}T_n\|$ alone give reason to believe that we are observing ion-frictional heating events sourced by soft cusp precipitation. However, we must bear in mind that for every F region ion heating event, this is accompanied by an E region ion heating of similar magnitude, even though the concurrent E region particle precipitation flux is weak (figure 7.2 a)). This demonstrates soft precipitation failing to steer the energy deposition exclusively to F region altitudes, as field-aligned currents in the E region are instigated, depositing Ohmic Joule heating. The E and F region ion temperature enhancements are of similar magnitudes; therefore, it is impossible to separate their effects on the upwellings.

After the cusp precipitation passes out of the FOV, both temperatures do not reduce immediately. The ion temperatures increase dramatically to ∼5000K, representative of the extreme heating event which occurs across all altitudes seen in figure 7.2d).The neutral temperatures remain high to the end of the dataset at 12UT. This could be due to the neutrals larger heat capacity or some solar EUV glancing over the horizon around noon.

Throughout the extreme event at 10:45-11:05UT, which occurs after the cusp precipitation passes out of the zenith FOV, the ion-neutral temperature difference rises to ∼4000K. Taking the neutral temperature to be 1200K, this corresponds to a relative ion-neutral velocity of 2490ms$^{-1}$. This is extreme and would be unlikely to be produced by anything other than ion-frictional heating. At the time of this event, the IMAGE magnetometer data show another large perturbation in the x-component of the geomagnetic field, a signature of currents flowing in the E region above Svalbard. This will result in some Ohmic Joule heating. At the F region, with the cusp precipitation passing out of the FOV, the local conductivity decreases and the ions are likely to be accelerated in a different direction. Evidence of this can be seen in the horizontal wind-fitted data where the zenith vector is seen to be directed north-eastward at 10:44UT and then abruptly turns north-westward at 10:51UT. This is likely due to the ions changing direction, and eventually dragging the neutrals in the same direction until they 'catch up', and after which the ion-frictional heating weakens at 11:05UT.

We can be confident that this event involves ion-frictional heating as it is exactly correlated with the electron temperature dramatically dropping below the ion temperature ($T_r$ <1). However, as mentioned for the previous upwellings, E and F region Joule heat-

ing occur simultaneously. The resulting upwelling is double-peaked reaching speeds of 100ms$^{-1}$. This is not an extraordinary upwelling compared to the previous upwellings which occur with cusp precipitation, although the ion temperatures are extreme. The ion temperatures are more intense and persistent, occurring for a full 15 minutes at 3000K, whereas the previous events would only last a few minutes at most. Moreover, the temperatures are uniform through the 180-220km height range whereas the previous temperatures were lower. This indicates that heating at this altitude range has little effect on lifting the atmosphere. This event also shows that cusp precipitation is not essential to drive perturbed neutral winds continually, it is just necessary to 'instigate' the mechanism which drives the perturbations initially.

The general trend in ion temperature correlates well with the upwellings indicating a causal relationship. It is interesting how the upwelling associated with the large heating event at 11UT is of an average magnitude, while a much weaker heating around 08:40UT is found to coincide with the largest upwelling. It is highly likely that this extreme heating event is due to a large horizontal ion shear increasing the ion-frictional drag. According to equation 7.1 at the ion temperature of 5000K and the neutral temperature of 1100K the $\|v_i\text{-}v_n\|$ value would equal ∼2500m/s.

## 7.3 Conclusion

Simultaneous observations of the ionosphere and thermosphere cusp region using the narrow field FPI and the ESR showed F region cusp precipitation concurrent with a sustained neutral upwelling lasting ∼3 hours on 22nd January 2012 from 08:24-11:18UT. The neutral wind responded immediately to the arrival of the cusp overhead, both in the horizontal and vertical direction. Throughout the period of upwelling, there were continuous 'pulses' of ion temperature increases with periods of 5-20 minutes reaching the E and F region. The final upwelling came after the cusp precipitation passed out of the FOV and the corresponding ion temperatures bridge the gap between the E and F region indicating the altitude region 180-220km has little effect on the upwelling.

We have reason to believe that our data shows soft particle precipitation associated with the cusp depositing its energy at high altitudes of 200-300km, ionising the region, actuating ion-frictional heating, with typical ion-neutral temperature differences of ∼2000K, corresponding to ion velocities of 1000ms$^{-1}$. Although we may not have full confirmation

that these are ion-frictional heating events, due to lack of horizontal plasma drift measurements, the magnitudes of temperature enhancement are difficult to explain in any other way. This is especially so for the extreme event after cusp precipitation, where the ion temperatures are substantially elevated above those of the electrons with an ion-neutral temperature difference of 4000K, corresponding to ion velocities of $\sim$2500ms$^{-1}$.

It is not possible to attribute the entire upwelling to the F region ion-frictional heating as there is always a corresponding occurrence of E region ion-heating; the Ohmic Joule heating is likely to be causing some atmospheric lift. It is most likely that Joule heating from both E (indirectly) and F (directly) regions are activated by the cusp precipitation and are contributing in concert to achieve the extreme upwellings observed.

# Chapter 8

# Conclusion

Observations of thermospheric winds, particularly vertical ones, have been presented in this thesis, measured from the Doppler shift of the atomic oxygen 630nm airglow emission (which peaks at 240km altitude), as seen by three Fabry-Perot Interferometers (FPIs) located in Scandinavia. One is situated at the auroral site, Kiruna, in Sweden. The other two are collocated at the polar cap site, Longyearbyen in Svalvard; one of which is a standard narrow-field FPI, similar to the one at Kiruna, and the other is a state-of-the-art wide field SCANning Doppler Imager (SCANDI). From all these data, emission intensities, line-of-sight (LOS) wind speeds and neutral temperatures are obtained.

SCANDI allows simultaneous measurements of many zones within a circular region of the sky with a diameter of 1076km; the field-of-view is split tangentially into annuli and then radially into zones. At the onset of solar maximum, SCANDI's map of the sky is increased in resolution from a 25 zone map to 61 and 91 zone maps, with an equal radial separation between SCANDI's annuli. SCANDI's existing infrastructure is updated and the calibration methods and data analysis are presented. I calculated the new geometry for the zone partitioning, mapped from the sky to the detector chip, and I updated the trigonometry used in the data analysis routines. I performed some data calibrations. I discovered a ring offset within the intensities and explored methods to remove it. The Herrero and Meriwether (1980) technique gives the best correction. I discovered SCANDI's zonally dependent baseline for wind speed observations is a mixture of an instrument function and a background wind.

I presented the algorithm to fit a horizontal wind-field to SCANDI's LOS wind map which I altered from the original to suit the new, higher-spatial-resolution, zone configurations. I verified the wind-fitting routine by producing and cross-comparing climatologies of the neutral wind field and of the SuperDARN ion drift field around the magnetic pole. I found the Svalbard wind field complies with my expectations of thermospheric dynamics and the wind-fit algorithm is justified. This study demonstrated an interplay between ion drag and the pressure gradient forces on the polar neutral atmosphere. Fine spatial structure and divergences within the wind field were correlated with that in the ion drift fields. Ion drag was shown to have a more dominant effect than expected.

A statistical analysis of the vertical winds from 2002-2009 at Kiruna and Svalbard is presented. First a study of the auroral oval's relationship to the two observing sites is presented where I compared two auroral oval models which show the same result: Svalbard is more likely to observe vertical winds sourced by auroral energy deposition at magnetic noon, whereas Kiruna is more likely to observe such winds at magnetic midnight. I confirmed Svalbard is poleward of the auroral oval at midnight from 2002-2009 by presenting the average diurnal emission intensity variation for Kiruna and Svalbard. I calculated the diurnal vertical wind variation for quiet and active geomagnetic activity and compared with the variations from the 1980s (Aruliah et al, 1991; Rees et al, 1984a). There are inconsistencies between the two sites and their relative position to the auroral oval; the Kiruna data seem reasonable, however there are many peculiarities in the Svalbard dataset, e.g., the further away from the auroral oval, the larger the vertical winds. I then present the data as frequency distributions, showing both sites' vertical winds have leptokurtic distributions and Svalbard's is offset from zero and heavily skewed. Evident, in these distributions is an excess of extreme upward vertical winds in the Svalbard positive tail, particularly in the midnight magnetic time sector. I proposed that these winds might be termed 'black swan events' with a generation mechanism not yet identified. This study poses a problem that is solved in the next chapter.

I present an investigation into whether Svalbard's 'black swan events' are real vertical winds or whether they are due to spectral contamination; either instrumental or from an additional contaminating line. I present six black swan event case studies from December 2007 as vertical wind, temperature and intensity signals. This demonstrates a one-to-one correspondence between the temperature and vertical wind and an inverse relationship with the intensity signal. The extreme vertical wind raw spectra are presented showing

them to have poor signal-to-noise ratio (SNR).

I discuss the difficulty in physically generating upwellings in the thermosphere. Using off-zenith look direction FPI data, I showed that all 'black swan events' are accompanied by divergences in the horizontal wind at low intensities. The presence of both upwelling and divergences is consistent with mass continuity. However, upticks in the vertical wind and temperature distributions at low SNR indicate spectral contamination that becomes apparent only when the 630nm emission rate is weak. I conclude that the extreme winds are false, not black swan events and that spectral contamination is responsible.

I investigated the hydroxyl (OH) line contamination. I simulated the 630nm line with and without the hydroxyl lines and show a trend in wind and temperature enhancement with 630nm emission rate. I ascertained a threshold emission rate value, below which the wind and temperature data are enhanced more than their population standard deviations: 40R for the wind values and 10R for the temperatures. The FPI data with <40R emission rate are removed and the data are re-analysed. I presented the cleaned data as a histogram which is symmetric and centred around zero; no longer positively skewed. This supports the hypothesis that hydroxyl is the cause of the Svalbard extreme upwelling values ('black swan events'). I formulated a new annual procedure in order to maintain a 'clean' database involving annual calibrations of the FPI's relative intensities with the MSP data. A useful further development would be to characterise the OH line accurately enough so as to remove it from the spectra and correct the contaminated data.

The Burnside relationship (Burnside et al, 1982) derivation was shown explicitly, i.e., a positive relationship between the vertical wind and the horizontal divergence with a constant of proportionality equal to the atmospheric scale height. I performed a statistical study of the relationship between the vertical and horizontal components of wind with the aim of discovering the Burnside relation in the data; hence an alternative method to calculate the vertical wind. Data from 2008-2014 was used. The FPI zenith measurement gave the vertical wind and SCANDI's off-zenith measurements are used to calculate the divergence. A strong positive relationship was found in 39% of the data and a large spread in the effective atmospheric scale heights which shows an average value of 281km; 5.6-9.4 times larger than expectations. The Burnside relation was first demonstrated successfully for a low latitude site. However, my results demonstrate that the Burnside relation is a not a suitable model for our high-latitude site. I found the assumptions of constant temperature and wind speed with altitude could contribute to the discrepency between the

Burnside model and our results, however, it is unlikely to be due to the assumption of constant geopotential height (as the maximum movement of a pressure level was found to be within the vertical wind error bars on the CMAT2 model). Future developments required are, firstly, a tristatic study of the three-dimensional wind vectors so the Burnside relation may be tested without any measurement ambiguity; and, secondly, a similar investigation for an FPI dataset at lower latitudes. This study also highlighted the need for an in depth investigation of hydrostatic equilibrium using our FPI database. This would involve a comparison between upwellings and temperature variations; an investigation which is only possible now the Svalbard FPI database can be cleaned of hydroxyl contamination.

The CUSPN campaign, led by the Atmospheric Physics Laboratory (APL), is presented showing the first simultaneous measurements of the charged and neutral cusp region by the EISCAT Svalbard Radar, the Svalbard FPI and SCANDI. These are supplemented by data from the colocated Meridian Scanning Photometer (MSP) and the IMAGE magnetometer network. The arrival of a CME and the cusp overhead is shown as deflections in the IMAGE data. I present the main datasets showing several extreme cusp upwellings persisting for $\sim$3 hours starting at the onset of electron particle precipitation overhead. These are correlated with ion and neutral temperature enhancements. I use the MSP data and SCANDI intensity sky-map plots data to locate the cusp overhead. The SCANDI wind fields show a slow response to the arrival of the CME, and cusp precipitation, re-ordering with a time lag of 20-50minutes. I used the ion-neutral temperature difference together with the electron-neutral temperature ratio as a proxy indicator for ion-frictional heating. I showed compelling, however not conclusive, evidence that ion-frictional heating within the F region is driving the extreme upwellings and feeding the density bulge. I also showed evidence to suggest ion-frictional heating is occurring at the E and F region simultaneously towards the end of the night. There are two more cusp case-studies which require analysis as above. These experiments made use of the newly fixed 32m ESR dish which enabled measurements of the horizontal plasma drifts. From these a Joule heating altitude profile can be produced and the altitude of ion-frictional heating can be conclusively identified.

In conclusion, given the UCL network of highly sensitive FPIs as tools, I focused on the thermospheric vertical winds which are a key aspect of thermospheric dynamics. I approached the topic of vertical wind from three different angles: a statistical analysis to characterise the winds; the dependency of the vertical winds on the horizontal divergence to test the validity of the Burnside relation; and finally, a case study of the cusp upwellings

to determine their generation mechanism. I have given an overview of the UCL FPI instruments, presenting my calibrations and verifications of the SCANDI database. In particular I found a distortion of the Svalbard vertical wind frequency distribution that is due to hydroxyl contamination. The Kiruna database needs to be fully investigated for the presence of OH so a cross-comparison between the two sites can be performed, definitely characterising the auroral and polar cap vertical winds. The bulk of the vertical wind data are found to be within $\pm 10 \mathrm{ms}^{-1}$ at Svalbard and roughly $\pm 20 \mathrm{ms}^{-1}$ in Kiruna; small enough for hydrostatic equilibrium to be a good assumption for modellers most of the time. Occasionally, however, there are extreme vertical wind events which are likely to violate hydrostatic equilibrium, such as the ones observed in the CUSPN campaign; when extreme cusp upwellings occur at the time of a CME arrival. The cusp density bulge generation mechanism is not definitively identified, although there is compelling evidence of soft particle precipitation steering energy deposition via ion-frictional heating to high altitudes where the atmosphere is tenuous enough to be lifted with relatively little energy. This thesis has demonstrated that there is much more to be learnt about vertical winds, now that accurate measurements are available. Hence the fuller observation of thermospheric wind systems must be further pursued.

This page was intentionally left blank

# Bibliography


[1] C. Anderson, M. Conde, and M. G. McHarg. Neutral thermospheric dynamics observed with two scanning Doppler imagers: 2. Vertical winds. *Journal of Geophysical Research (Space Physics)*, 117:3305, March 2012.

[2] C. Anderson, M. Conde, and M. G. McHarg. Neutral thermospheric dynamics observed with two scanning Doppler imagers: 3. Horizontal wind gradients. *Journal of Geophysical Research (Space Physics)*, 117:5311, May 2012.

[3] C. Anderson, I. Yiu, A. C. Kellerman, M. Kosch, R. Makarevitch, A. L. Aruliah, M. G. Conde, T. Davies, and P. L. Dyson. First Observations of Simultaneous Interhemispheric Conjugate High-latitude Thermospheric Winds. *AGU Fall Meeting Abstracts*, page A1459, December 2009.

[4] Callum Anderson. *Remote Optical Sensing of Local-Scale Thermospheric Dynamics Above Antarctica*. PhD thesis, La Trobe University, 2011.

[5] Callum Anderson, Theo Davies, Mark Conde, Peter Dyson, and MJ Kosch. Spatial sampling of the thermospheric vertical wind field at auroral latitudes. *Journal of Geophysical Research: Space Physics (1978–2012)*, 116(A6), 2011.

[6] A. Aruliah and M. Kosch. Determining the 630nm emission altitude using modelling and observations from a tristatic configuration of Fabry-Perot Interferometers and EISCAT radars. In *38th COSPAR Scientific Assembly*, volume 38 of *COSPAR Meeting*, page 1144, 2010.

[7] A. L. Aruliah, E. M. Griffin, I. McWhirter, A. D. Aylward, E. A. K. Ford, A. Charalambous, M. J. Kosch, C. J. Davis, and V. S. C. Howells. First tristatic studies of meso-scale ion-neutral dynamics and energetics in the high-latitude upper atmosphere using collocated FPIs and EISCAT radar. *GRL*, 31:3802, February 2004.

[8] A. L. Aruliah and D. Rees. The trouble with thermospheric vertical winds: Geomagnetic, seasonal and solar cycle dependence at high latitudes. *Journal of Atmospheric and Terrestrial Physics*, 57:597–609, May 1995.

[9] AL Aruliah and E Griffin. Evidence of meso-scale structure in the high-latitude thermosphere. In *Annales Geophysicae*, volume 19, pages 37–46, 2001.

[10] AL Aruliah, EM Griffin, H-CI Yiu, I McWhirter, and A Charalambous. Scandi–an all-sky doppler imager for studies of thermospheric spatial structure. In *Annales Geophysicae*, volume 28, pages 549–567. Copernicus GmbH, 2010.

[11] Anasuya Aruliah. *The Synoptic Variability of Thermospheric and Mesospheric Winds Observed Using a Fabry Perot Interferometer*. PhD thesis, University College London, 1991.

[12] Anasuya L Aruliah, Ingo CF Müller-Wodarg, and Jacqueline Schoendorf. Consequences of geomagnetic history on the high-latitude thermosphere and ionosphere: Averages. *Journal of Geophysical Research: Space Physics (1978–2012)*, 104(A12):28073–28088, 1999.

[13] Harold D Babcock. A study of the green auroral line by the interference method. *The Astrophysical Journal*, 57:209, 1923.

[14] Manfred A Biondi and Dwight P Sipler. Horizontal and vertical winds and temperatures in the equatorial thermosphere: Measurements from natal, brazil during august–september 1982. *Planetary and space science*, 33(7):817–823, 1985.

[15] Asgeir Brekke. *Physics of the upper polar atmosphere*. Springer Science & Business Media, 2012.

[16] KA Browning and R Wexler. The determination of kinematic properties of a wind field using doppler radar. *Journal of Applied Meteorology*, 7(1):105–113, 1968.

[17] RG Burnside, RA Behnke, and JCG Walker. Meridional neutral winds in the thermosphere at arecibo: Simultaneous incoherent scatter and airglow observations. *Journal of Geophysical Research: Space Physics (1978–2012)*, 88(A4):3181–3189, 1983.

[18] Herbert C Carlson, Todd Pedersen, Santimay Basu, Mike Keskinen, and Jøran Moen.

Case for a new process, not mechanism, for cusp irregularity production. *Journal of Geophysical Research: Space Physics (1978–2012)*, 112(A11), 2007.

[19] Herbert C Carlson, Timothy Spain, Anasuya Aruliah, Asmund Skjaeveland, and Joran Moen. First-principles physics of cusp/polar cap thermospheric disturbances. *Geophysical Research Letters*, 39(19), 2012.

[20] Joseph W Chamberlain. Spectral photometry of the nightglow. *Physics of the Aurora and Airglow*, pages 486–520, 1961.

[21] Joseph W Chamberlain et al. Physics of the aurora and airglow. 1961.

[22] JH Clemmons, JH Hecht, DR Salem, and DJ Strickland. Thermospheric density in the earth's magnetic cusp as observed by the streak mission. *Geophysical Research Letters*, 35(24), 2008.

[23] M. V. Codrescu, T. J. Fuller-Rowell, and J. C. Foster. On the importance of E-field variability for Joule heating in the high-latitude thermosphere. *GRL*, 22:2393–2396, 1995.

[24] M. Conde, J. D. Craven, T. Immel, E. Hoch, H. Stenbaek-Nielsen, T. Hallinan, R. W. Smith, J. Olson, W. Sun, L. A. Frank, and J. Sigwarth. Assimilated observations of thermospheric winds, the aurora, and ionospheric currents over Alaska. *JGR*, 106:10493–10508, June 2001.

[25] M. Conde and P. L. Dyson. Thermospheric vertical winds above Mawson, Antarctica. *Journal of Atmospheric and Terrestrial Physics*, 57:589–596, May 1995.

[26] M. Conde and R. W. Smith. Spatial structure in the thermospheric horizontal wind above Poker Flat, Alaska, during solar minimum. *JGR*, 103:9449–9472, May 1998.

[27] Robert R Conway, Michael H Stevens, Joel G Cardon, Scott E Zasadil, Charles M Brown, Jeff S Morrill, and George H Mount. Satellite measurements of hydroxyl in the mesosphere. *Geophysical research letters*, 23(16):2093–2096, 1996.

[28] SL Cooper, M Conde, and P Dyson. Numerical simulations of thermospheric dynamics: divergence as a proxy for vertical winds. In *Annales geophysicae: atmospheres, hydrospheres and space sciences*, volume 27, page 2491, 2009.

[29] Russell B Cosgrove, Hasan Bahcivan, Steven Chen, Robert J Strangeway, Juan Ortega, Mohamed Alhassan, Yuchen Xu, Megan Van Welie, James Rehberger, Sebastian Musielak, et al. Empirical model of poynting flux derived from fast data and a cusp signature. *Journal of Geophysical Research: Space Physics*, 119(1):411–430, 2014.

[30] R. I. Crickmore. A comparison between vertical winds and divergence in the high-latitude thermosphere. *Annales Geophysicae*, 11:728–733, August 1993.

[31] A Damiani, M Storini, C Rafanelli, and P Diego. The hydroxyl radical as an indicator of sep fluxes in the high-latitude terrestrial atmosphere. *Advances in Space Research*, 46(9):1225–1235, 2010.

[32] HG Demars and RW Schunk. Thermospheric response to ion heating in the dayside cusp. *Journal of atmospheric and solar-terrestrial physics*, 69(6):649–660, 2007.

[33] Y. Deng, T. J. Fuller-Rowell, A. J. Ridley, D. Knipp, and R. E. Lopez. Theoretical study: Influence of different energy sources on the cusp neutral density enhancement. *Journal of Geophysical Research (Space Physics)*, 118:2340–2349, May 2013.

[34] Yue Deng and Aaron J Ridley. Possible reasons for underestimating joule heating in global models: E field variability, spatial resolution, and vertical velocity. *Journal of Geophysical Research: Space Physics (1978–2012)*, 112(A9), 2007.

[35] Margit Dyrland. *Multi-instrument studies of polar mesopause region temperature and airglow variability*. PhD thesis, University Centre in Svalbard, 2010.

[36] PJ Espy and J Stegman. Trends and variability of mesospheric temperature at high-latitudes. *Physics and Chemistry of the Earth, Parts A/B/C*, 27(6):543–553, 2002.

[37] D. S. Evans. The observations of a near monoenergetic flux of auroral electrons. *JGR*, 73:2315, April 1968.

[38] D. S. Evans. Precipitating electron fluxes formed by a magnetic field aligned potential difference. *JGR*, 79:2853, 1974.

[39] Ch Fabry and Alfred Perot. On a new form of interferometer. *The Astrophysical Journal*, 13:265, 1901.

[40] Ya I Feldstein and GV Starkov. Auroral oval in the igy and iqsy period and a ring current in the magnetosphere. *Planetary and Space Science*, 16(1):129–133, 1968.

[41] EAK Ford, AL Aruliah, EM Griffin, and I McWhirter. High time resolution measurements of the thermosphere from fabry-perot interferometer measurements of atomic oxygen. In *Annales Geophysicae*, volume 25, pages 1269–1278. Copernicus GmbH, 2007.

[42] JT Gosling. Variations in the solar wind speed along the earth's orbit. *Solar Physics*, 17(2):499–508, 1971.

[43] P. A. Greet, J. L. Innis, and P. L. Dyson. Thermospheric vertical winds in the auroral oval/polar cap region. *Annales Geophysicae*, 20:1987–2001, December 2002.

[44] W Guo and DJ McEwen. Vertical winds in the central polar cap. *Geophysical research letters*, 30(14), 2003.

[45] Dø A Hardy, MS Gussenhoven, R Raistrick, and WJ McNeil. Statistical and functional representations of the pattern of auroral energy flux, number flux, and conductivity. *Journal of Geophysical Research: Space Physics (1978–2012)*, 92(A11):12275–12294, 1987.

[46] J. K. Hargreaves, H. J. A. Chivers, and E. Nielsen. Properties of spike events in auroral radio absorption. *JGR*, 84:4245–4250, August 1979.

[47] John K Hargreaves. *The solar-terrestrial environment*. Cambridge Press, 1992.

[48] R. A. Heelis, D. McEwen, and W. Guo. Ion and neutral motions observed in the winter polar upper atmosphere. *Journal of Geophysical Research (Space Physics)*, 107:1476, December 2002.

[49] G. Hernandez. Contamination of the O I ( $^3$P $_2$- $^1$D $_2$) emission line by the (9-3) band of OH X $^2$II in high-resolution measurements of the night sky. *JGR*, 79:1119, 1974.

[50] G. Hernandez and R. G. Roble. Thermospheric nighttime neutral temperature and winds over Fritz Peak Observatory: Observed and calculated solar cycle variation. *JGR*, 100:14647–14660, August 1995.

[51] FA Herrero and JW Meriwether. 6300-å airglow meridional intensity gradients. *Journal of Geophysical Research: Space Physics (1978–2012)*, 85(A8):4191–4204, 1980.

[52] JM Holmes, M Conde, C Deehr, and D Lummerzheim. Morphology of evening sector aurorae in $\lambda$557. 7-nm doppler temperatures. *Geophysical research letters*, 32(2), 2005.

[53] AJ Hundhausen. The solar wind. *Introduction to space physics*, pages 91–128, 1995.

[54] J. L. Innis and M. Conde. High-latitude thermospheric vertical wind activity from Dynamics Explorer 2 Wind and Temperature Spectrometer observations: Indications of a source region for polar cap gravity waves. *Journal of Geophysical Research (Space Physics)*, 107:1172, August 2002.

[55] J. L. Innis, P. A. Greet, and P. L. Dyson. Fabry-Perot spectrometer observations of the auroral oval/polar cap boundary above Mawson, Antarctica. *Journal of Atmospheric and Terrestrial Physics*, 58:1973–1988, December 1996.

[56] JL Innis and M Conde. Thermospheric vertical wind activity maps derived from dynamics explorer-2 wats observations. *Geophysical research letters*, 28(20):3847–3850, 2001.

[57] M. Ishii, M. Conde, R. W. Smith, M. Krynicki, E. Sagawa, and S. Watari. Vertical wind observations with two Fabry-Perot interferometers at Poker Flat, Alaska. *JGR*, 106:10537–10552, June 2001.

[58] TL Killeen. Energetics and dynamics of the earths thermosphere. *Reviews of Geophysics*, 25(3):433–454, 1987.

[59] D. J. Knipp, W. K. Tobiska, and B. A. Emery. Direct and Indirect Thermospheric Heating Sources for Solar Cycles 21-23. *Solar Physics*, 224:495–505, October 2004.

[60] M. J. Kosch, K. Cierpka, M. T. Rietveld, T. Hagfors, and K. Schlegel. High-latitude ground-based observations of the thermospheric ion-drag time constant. *GRL*, 28:1395–1398, April 2001.

[61] M. J. Kosch, M. Ishii, A. Kohsiek, D. Rees, K. Schlegel, T. Hagfors, and K. Cierpka. A Comparison of Vertical Thermospheric Winds from Fabry-Perot Interferometer Measurements over a 50 km Baseline. *Advances in Space Research*, 26:985–988, 2000.

[62] M. J. Kosch, A. Senior, C. Bryers, M. T. Rietveld, B. Xu, J. Wu, and T. Pedersen. Unusual ionospheric plasma density enhancements during high-power ionospheric pumping at EISCAT. *AGU Fall Meeting Abstracts*, page A7, December 2012.

[63] M. J. Kosch, I. Yiu, C. Anderson, T. Tsuda, Y. Ogawa, S. Nozawa, A. Aruliah, V. Howells, L. J. Baddeley, I. W. McCrea, and J. A. Wild. Mesoscale observations

of Joule heating near an auroral arc and ion-neutral collision frequency in the polar cap E region. *Journal of Geophysical Research (Space Physics)*, 116:5321, May 2011.

[64] J. Kurihara, S. Oyama, S. Nozawa, T. T. Tsuda, R. Fujii, Y. Ogawa, H. Miyaoka, N. Iwagami, T. Abe, K.-I. Oyama, M. J. Kosch, A. Aruliah, E. Griffin, and K. Kauristie. Temperature enhancements and vertical winds in the lower thermosphere associated with auroral heating during the DELTA campaign. *Journal of Geophysical Research (Space Physics)*, 114:12306, December 2009.

[65] M. F. Larsen and J. W. Meriwether. Vertical winds in the thermosphere. *Journal of Geophysical Research (Space Physics)*, 117:9319, September 2012.

[66] H. Lühr, M. Rother, W. Köhler, P. Ritter, and L. Grunwaldt. Thermospheric upwelling in the cusp region: Evidence from CHAMP observations. *GRL*, 31:6805, March 2004.

[67] LR Lyons, TL Killeen, and RL Walterscheid. The neutral wind flywheel as a source of quiet-time, polar-cap currents. *Geophysical research letters*, 12(2):101–104, 1985.

[68] K Minschwaner, GL Manney, NJ Livesey, HC Pumphrey, HM Pickett, L Froidevaux, A Lambert, MJ Schwartz, PF Bernath, KA Walker, et al. The photochemistry of carbon monoxide in the stratosphere and mesosphere evaluated from observations by the microwave limb sounder on the aura satellite. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 115(D13), 2010.

[69] A. Muellemann and F.-J. Luebken. Horizontal winds in the mesosphere at high latitudes. In J.-P. Paillé, editor, *35th COSPAR Scientific Assembly*, volume 35 of *COSPAR Meeting*, page 2035, 2004.

[70] P. T. Newell, T. Sotirelis, and S. Wing. Diffuse, monoenergetic, and broadband aurora: The global precipitation budget. *Journal of Geophysical Research (Space Physics)*, 114:9207, September 2009.

[71] RJ Niciejewski, Timothy L Killeen, and Y Won. Observations of neutral winds in the polar cap during northward imf. *Journal of atmospheric and terrestrial physics*, 56(2):285–295, 1994.

[72] S. Oyama, B. J. Watkins, S. Nozawa, S. Maeda, and J. Watermann. Initial Results Of The Lower Thermospheric Vertical Wind Estimated With Incoherent-Scatter Radars At High Latitudes. *AGU Fall Meeting Abstracts*, page A1105, December 2004.

[73] D. Rees, P. J. Charleton, N. Lloyd, R. W. Smith, F. G. McCormac, and A. Steen. The generation of vertical thermospheric winds and gravity waves at auroral latitudes. I - Observations of vertical winds. *Planetary Space Science*, 32:667–684, June 1984.

[74] M. H. Rees. *Physics and Chemistry of the Upper Atmosphere*. August 1989.

[75] MH Rees and RG Roble. Excitation of o (1 d) atoms in aurorae and emission of the [oi] 6300-å line. *Canadian journal of physics*, 64(12):1608–1613, 1986.

[76] Henry Rishbeth and Owen K Garriott. Introduction to ionospheric physics. *IEEE Transactions on Image Processing*, 1, 1969.

[77] R. G. Roble, E. C. Ridley, and R. E. Dickinson. On the global mean structure of the thermosphere. *JGR*, 92:8745–8758, August 1987.

[78] RG Roble and BA Emery. On the global mean temperature of the thermosphere. *Planetary and Space Science*, 31(6):597–614, 1983.

[79] AS Rodger, GD Wells, RJ Moffett, and GJ Bailey. The variability of joule heating, and its effects on the ionosphere and thermosphere. In *Annales Geophysicae*, volume 19, pages 773–781, 2001.

[80] F. B. Sadler, M. Lessard, E. Lund, A. Otto, and H. Lühr. Auroral precipitation/ion upwelling as a driver of neutral density enhancement in the cusp. *Journal of Atmospheric and Solar-Terrestrial Physics*, 87:82–90, October 2012.

[81] Kristian Schlegel, Hermann Lühr, J-P St-Maurice, Geoff Crowley, and Chris Hackert. Thermospheric density structures over the polar regions observed with champ. In *Annales Geophysicae*, volume 23, pages 1659–1672. Copernicus GmbH, 2005.

[82] Donald E Shemansky and A Vallance Jones. Type-b red aurora; the o 2+ first negative system and the n 2 first positive system. *Planetary and Space Science*, 16(9):1115–1130, 1968.

[83] MF Smith and Mike Lockwood. Earth's magnetospheric cusps. *Reviews of Geophysics*, 34(2):233–260, 1996.

[84] R. W. Smith. Vertical winds: a tutorial. *Journal of Atmospheric and Solar-Terrestrial Physics*, 60:1425–1434, September 1998.

[85] R. W. Smith and G. Hernandez. Vertical winds in the thermosphere within the polar cap. *Journal of Atmospheric and Terrestrial Physics*, 57:611–620, May 1995.

[86] S. C. Solomon, P. B. Hays, and V. J. Abreu. The auroral 6300 A emission - Observations and modeling. *JGR*, 93:9867–9882, September 1988.

[87] N. W. Spencer, L. E. Wharton, G. R. Carignan, and J. C. Maurer. Thermosphere zonal winds - Vertical motions and temperature as measured from Dynamics Explorer. *GRL*, 9:953–956, September 1982.

[88] J-P St-Maurice and WB Hanson. Ion frictional heating at high latitudes and its possible use for an in situ determination of neutral thermospheric winds and temperatures. *Journal of Geophysical Research: Space Physics (1978–2012)*, 87(A9):7580–7602, 1982.

[89] KF Tapping. Recent solar radio astronomy at centimeter wavelengths: The temporal variability of the 10.7-cm flux. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 92(D1):829–838, 1987.

[90] J. P. Thayer. Height-resolved Joule heating rates in the high-latitude E region and the influence of neutral winds. *JGR*, 103:471–487, January 1998.

[91] J. P. Thayer and J. Semeter. The convergence of magnetospheric energy flux in the polar atmosphere. *Journal of Atmospheric and Solar-Terrestrial Physics*, 66:807–824, July 2004.

[92] M. R. Torr, D. G. Torr, and P. G. Richards. A new determination of the ultraviolet heating efficiency of the thermosphere. *JGR*, 85:6819–6826, December 1980.

[93] A Vallance Jones. The telluric emission spectrum in the range 1 $\mu$ to 3 mm: Introductory report. *Mem. Soc. Roy. Sci. Liege, Ser*, 5(9):289, 1964.

[94] James F Vickrey, Richard R Vondrak, and Stephen J Matthews. Energy deposition by precipitating particles and joule dissipation in the auroral ionosphere. *Journal of Geophysical Research: Space Physics (1978–2012)*, 87(A7):5184–5196, 1982.

[95] Michael N Vlasov, Michael J Nicolls, Michael C Kelley, Steven M Smith, Néstor Aponte, and Sixto A González. Modeling of airglow and ionospheric parameters at arecibo during quiet and disturbed periods in october 2002. *Journal of Geophysical Research: Space Physics (1978–2012)*, 110(A7), 2005.

[96] Iris Yiu. *High latitude thermosphere mesoscale studies and long-term database investigations with the new Scanning Doppler Imager and Fabry-Perot Interferometers.* PhD thesis, University College London, 2011.

[97] B. Zhang, W. Lotko, O. Brambles, M. J. Wiltberger, W. Wang, P. Schmitt, and J. Lyon. Effects of Soft Electron Precipitation on the Coupled Magnetosphere-Ionosphere-Thermosphere. *AGU Fall Meeting Abstracts*, page A2169, December 2012.